

## Deciphering the Molecular Diversity of an Ant Venom Peptidome through a Venomics Approach

Axel Touchard,<sup>\*,†,ID</sup> Nathan Téné,<sup>†</sup> Philippe Chan Tchi Song,<sup>‡</sup> Benjamin Lefranc,<sup>§</sup> Jérôme Leprince,<sup>§,ID</sup> Michel Treilhou,<sup>†,#</sup> and Elsa Bonnafé<sup>†,#</sup>

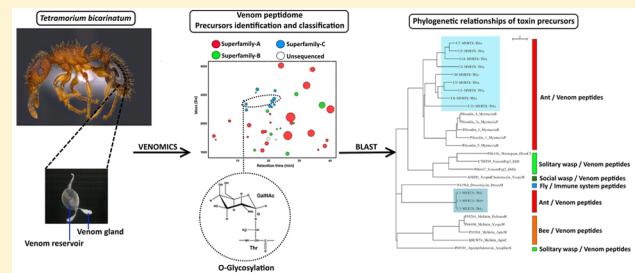
<sup>†</sup>Equipe BTSB-EA 7417, Université de Toulouse, Institut National Universitaire Jean-François Champollion, Place de Verdun, 81012 Albi, France

<sup>‡</sup>Normandie Univ, UNIROUEN, Institut de Recherche et d'Innovation Biomédicale (IRIB), 76000 Rouen, France

<sup>§</sup>Inserm U 1239, Normandie Univ, UNIROUEN, Plate-forme de Recherche en Imagerie Cellulaire Normandie (PRIMACEN), 76000 Rouen, France

### Supporting Information

**ABSTRACT:** The peptide toxins in the venoms of small invertebrates such as stinging ants have rarely been studied due to the limited amount of venom available per individual. We used a venomics strategy to identify the molecular diversity of the venom peptidome for the myrmicine ant *Tetramorium bicarinatum*. The methodology included (i) peptidomics, in which the venom peptides are sequenced through a *de novo* mass spectrometry approach or Edman degradation; (ii) transcriptomics, based on RT-PCR-cloning and DNA sequencing; and (iii) the data mining of the RNA-seq in the available transcriptome. Mass spectrometry analysis revealed about 2800 peptides in the venom. However, the *de novo* sequencing suggested that most of these peptides arose from processing or the artifactual fragmentations of full-length mature peptides. These peptides, called “myrmicitoxins”, are produced by a limited number of genes. Thirty-seven peptide precursors were identified and classified into three superfamilies. These precursors are related to pilosulin, secapin or are new ant venom prepropeptides. The mature myrmicitoxins display sequence homologies with antimicrobial, cytolytic and neurotoxic peptides. The venomics strategy enabled several post-translational modifications in some peptides such as O-glycosylation to be identified. This study provides novel insights into the molecular diversity and evolution of ant venoms.



**KEYWORDS:** *Tetramorium bicarinatum*, ant venom, peptidome, transcriptomics, peptidomics, venomics, peptides, post-translational modification, O-glycosylation

## 1. INTRODUCTION

Venoms are sophisticated biochemical cocktails comprising a large array of bioactive compounds.<sup>1</sup> Peptides are the predominant component of most arthropod venoms and are currently under investigation for their potential as therapeutic lead compounds, molecular probes or insect-selective biopesticides.<sup>2,3</sup>

In contrast to cone snails, snakes, scorpions and spiders, the venom composition of a wide range of venomous insects has been less studied.<sup>4</sup> The study of such neglected organisms is important toward understanding the vast diversity of venomous systems in the animal kingdom and providing novel insights into the evolution of venom toxins. Ants (Hymenoptera: Formicidae) belong to a dominant, diverse but also neglected group of venomous arthropods.<sup>5</sup> Although mass spectrometry-based studies of a variety of ant venoms have revealed the presence of thousands of peptides per species some of which possess novel structural features,<sup>6,7</sup> less than 100 venom peptides have been characterized from a limited number of ant species.<sup>8</sup> The scarcity of studies on ant venom

peptides is due to the small size of the specimens and the difficulties related to collecting their venoms. However, the advances in “omics” technologies now allow researchers to overcome these limitations and explore such venom compositions in detail. Recent transcriptomic studies (RNA-Seq) of venom glands have started to uncover the genes encoding the molecular diversity of the ant venom peptides.<sup>9–11</sup>

Transcriptomic-based studies, particularly those using next generation sequencing platforms, provide a highly sensitive and high-throughput technique for the exploration of venom composition at the RNA level. This approach yields a set of putative precursor sequences from which the mature peptides can subsequently be predicted. Nevertheless, the shortness of open reading frames (ORFs) encoding peptides complicates their identification, and predicting cleavage maturation sites can be difficult due to their nonconventional nature.

**Received:** June 13, 2018

**Published:** August 27, 2018

Furthermore, transcriptomic RNA-Seq analyses may generate numerous inherent errors within the transcriptome data set due to the *de novo* assembly of short reads into contigs leading to shifted or incomplete ORFs. Proteomic analyses, typically liquid chromatography coupled with mass spectrometry (LC–MS), are often used in combination with transcriptomic data and enable the direct examination of venom peptides in order to confirm and/or identify predicted sequences from transcriptomics data.<sup>12,13</sup> Additionally, mass spectrometry permits the identification of the post-translational modifications (PTMs) that are often observed in mature venom peptides. Such an integrated venomics approach has revolutionized the field of toxinology in well-studied venomous organisms generating an accurate picture of their venom composition.<sup>4</sup> Venomics studies have also begun to shed light on the molecular diversity of venom from a range of poorly studied venomous taxa and very recently have been applied to one ant venom.<sup>9</sup> Indeed, most of the previous studies of ant venoms were conducted on isolated peptides and there is currently no comprehensive investigation of the repertoire of peptide toxins found in the venom of a single species. The aim of this study was to uncover the range of molecular diversity of the peptides expressed in the venom glands of the ant *Tetramorium bicarinatum*.

The venom of *T. bicarinatum* is comprised of peptides and proteins and its two most abundant peptides, called “bicarinatin” and “P17”, were previously isolated and characterized from a set of 31 peptides detected by LC–MS.<sup>14</sup> At the same time, the venom gland transcriptome (RNA-Seq) of *T. bicarinatum* was published and provides a few new putative peptide and protein toxins.<sup>11</sup> In order to gain further insight into the venom peptide arsenal of *T. bicarinatum*, we used an integrated venomics approach that combines transcriptome shotgun assembly (TSA) with data from peptidomics (mass fingerprinting by LC–MS, *de novo* sequencing by LC–MS/MS or Edman degradation). Finally, we confirmed and completed cDNA sequences and the gene expression in the venom glands using a direct transcriptomic approach (RT-PCR based-cloning). We identified a total of 37 peptide toxin precursors. Further clustering analyses permitted us to classify these precursors into three superfamilies.

## 2. MATERIALS AND METHODS

### 2.1. Ants and Venom Sample Preparation

Two colonies of *T. bicarinatum* were collected by Dr. Jérôme Orivel in Itabuna, Bahia state, Brazil in 2000 and placed in our laboratory. These ant colonies are the same as those used by Bouzid et al. for the transcriptomic studies.<sup>11,15</sup> The colonies were maintained at 25 °C and fed three times a week with fresh mealworms and an aqueous honey solution (1:1 v/v). The ants’ venom sacs were dissected and placed together in water containing 1% formic acid (FA) (v/v) and the membranes disrupted using ultrasonic waves for 2 min. Then, the samples were centrifuged for 5 min at 14 400 rpm, the supernatant was collected and dried with a speed vacuum prior to storage at –20 °C until use. Venom samples from 100 venom sacs were mixed together for all of the proteomic analyses.

### 2.2. Mass Spectrometry Analysis and Venom Fractionation

A preliminary LC–MS analysis of the crude venom was carried out on the LCQ-Ion trap Surveyor equipped with an ESI-LC system Advantage (ThermoFisher Scientific, France). Peptides

were separated using a Luna-C18 column (5 μm; 2 × 150 mm; Phenomenex, France). The mobile phase was a gradient prepared from 0.1% aqueous formic acid (solvent A) and 0.1% formic acid in acetonitrile (solvent B). Peptides were eluted using a linear gradient from 0 to 50% B during 45 min, from 50% to 100% B during 10 min and then held for 5 min at a 200 μL min<sup>−1</sup> flow rate. The electrospray ionization mass spectrometry detection was performed in positive mode with the following optimized parameters: the capillary temperature was set at 300 °C, the spray voltage was 4.5 kV, the sheath gas and auxiliary gas were set at 50 and 10 psi, respectively. The acquisition range was from 100 to 2000 *m/z*. The fractions were collected each minute and dried for further Edman degradation. The area value for each peak corresponding to a peptide was manually integrated using the peak ion extraction function in Xcalibur software (ThermoFisher Scientific). The relative peak area indicates the contribution of each peptide to all of the peptides identified in the venom, providing a measure of relative abundance.

### 2.3. De Novo Orbitrap Mass Spectrometry-Based Sequencing

Venom fractions containing isolated peptides (i.e., Tb-2081, Tb-4018, Tb-2671 and Tb-3187) were incubated in 10 μL of ammonium bicarbonate buffer (pH 8) and 10 mM dithiothreitol (DTT) for 30 min at 56 °C. Then, the reduced fractions were alkylated by incubation in 50 mM iodoacetic acid (IAA) for 15 min at room temperature in the dark. An aliquot of each reduced/alkylated peptide was subjected to *de novo* sequencing while another aliquot was subjected to Edman degradation. Crude venom and fractions were subjected to *de novo* sequencing using an LTQ-Orbitrap Elite coupled to an Easy nLC II system (both from Thermo Fisher Scientific). Samples were injected into an enrichment column (C18 PepMap100, ThermoFisher Scientific). Separation was achieved using an analytical column needle (NTCC-360/100–5–153, NikkyoTechnos, Tokyo, Japan). The mobile phase consisted of H<sub>2</sub>O/FA 0.1% (buffer A) and ACN/FA 0.1% (buffer B). The peptides were eluted at a flow rate of 300 nL min<sup>−1</sup> using a linear gradient from 5 to 45% B during 109 min. The mass spectrometer was operated in positive ionization mode with capillary voltage and a source temperature set at 1.5 kV and 275 °C, respectively. The samples were analyzed using both higher-energy collision dissociation (HCD) and collision induced dissociation (CID) methods. The first scan (MS spectra) was recorded by the Orbitrap analyzer (*R* = 60 000) with a mass range of *m/z* 400–1800. Then, the 10 or 20 most intense ions were selected for MS2 experiments in HCD (resolution at 30 000) or CID modes, respectively. Only di- and tricharged ions were selected; other charge states were excluded for the MS2 experiments. The dynamic exclusion of already-fragmented precursor ions was applied for 30 s, with a repeat count of 1, a repeat duration of 30 s and an exclusion mass width of ±10 ppm. The precursor isolation width was 2 *m/z*. Fragmentation occurred in the linear ion trap analyzer with a normalized collision energy of 35%. All measurements using the Orbitrap analyzer were performed with on-the-fly internal recalibration (lock mass) at *m/z* 445.12002 (polydimethylcyclosiloxane). Peptide *de novo* sequencing was conducted using the PEAKS studio 7.5 proteomics workbench (Bioinformatics Solutions Inc., Waterloo, Canada) with the following specific parameters: enzyme, none; variable modifications, oxidation (M), amidation,

deamidation (NQ), pyro-glu from E and Q; monoisotopic; mass tolerance for precursor ions, 20 ppm; mass tolerance for fragment ions, 50 ppm; MS scan mode, Orbitrap; MS/MS scan mode, linear ion trap. The high confident PEAKS *de novo* sequencing result was filtered using the average local confidence score (ALC  $\geq$  70%). The mass spectrometry data were deposited in the ProteomeXchange Consortium via the PRIDE partner repository (<https://www.ebi.ac.uk/pride/archive/>) with the data set identifier PXD010451.

#### 2.4. Edman Sequencing

The primary structures of the peptides were determined through automated Edman degradation. Venom fractions containing isolated peptides were resuspended in an acetonitrile, trifluoroacetic acid, and water solution (30/1/69 v/v/v) and were loaded onto a precycled Biobrene Plus-coated glass filter. The N-terminal sequences were then determined by introducing the filter disc into a Procise P494 automated protein sequenator (Applied Biosystems, Foster City, CA) and carrying out runs of Edman degradation.

#### 2.5. Glycosylation Study

Synthetic Fmoc-Thr(O-GalNAc)-OH or O-GlcNAc (2 pmol) and crude venom were respectively analyzed with a nano-LC1200 system coupled to a Q-TOF 6550 mass spectrometer equipped with a nanospray source and an HPLC-chip cube interface (Agilent Technologies, Santa Clara, CA). For crude venom, a 26 min linear gradient (3–75% acetonitrile in 0.1% FA), at a flow rate of 350 nL min<sup>-1</sup>, was used to separate peptides onto a polaris-HR-Chip C18 column (150 mm long  $\times$  75  $\mu$ m inner diameter). Full autoMS1 scans from 290 to 1700 *m/z* and autoMS2 from 59 to 1700 *m/z* were recorded. For each cycle, a maximum of five precursors were isolated and fragmented into the collision cell with a fixed collision energy (i.e., 20, 25, 27, 30, and 35 eV). Each experiment was set to use an inclusion list for the single charge of the synthetic O-Thr and the 2+ and 3+ of each specific peptide. The active exclusion of these precursors was enabled after one spectrum prior to 0.13 min, and the absolute threshold for precursor selection was set at 1000 counts (relative threshold 0.001%).

#### 2.6. Direct Sequencing of Predicted cDNA

To confirm or to complete cDNA sequences and to verify that the genes encoding the peptides are expressed in venom glands, we isolated cDNA encoding peptides through a direct molecular approach. Briefly, all of the RNAs from 200 venom glands were extracted in a TRIzol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. Contaminating genomic DNA was removed using a DNA-free kit (Applied Biosystem) by following the manufacturer's instructions. RNA quantity was evaluated using a nanodrop (Nanodrop 2000, ThermoFisher Scientific). To confirm complete cDNA sequences, cDNA was synthesized from 100 ng of the total RNA using a M-MLV reverse transcriptase (Invitrogen) at 37 °C for 50 min with random hexamers according to the protocol of the manufacturer. To complete the cDNA sequences, cDNA was synthesized from 500 ng of total RNA with M-MLV (Invitrogen) and a 5'RACE adapter from the FirstChoice RLM-RACE kit (Invitrogen) by following the protocol of the manufacturer. PCR experiments were performed by using specific reverse primers designed manually with predicted sequences and 5' inner and outer adapter primers from the FirstChoice RLM-RACE kit in order to obtain the complete 5' cDNA part of sequence encoding the

prepro-peptide. Then, the PCR products were either cleaned using the PureLink PCR purification kit (Invitrogen) or purified using agarose gel when several products were amplified thanks to the PureLink Quick Gel Extraction Kit (Invitrogen) and sequenced on a Get-TQ platform (Purpan, Toulouse, France). Complete ORFs were determined using a translate program and subsequently aligned using the Seaview platform with the muscle program and formatted using BOXSHADE 3.3.1 program.

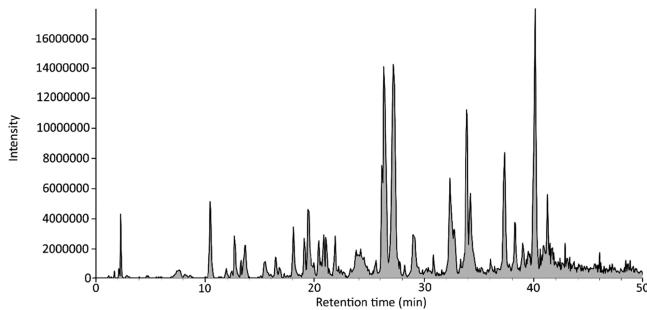
#### 2.7. Bioinformatic Tools

Peptide sequences obtained through mass spectrometry or prepro-peptide sequences obtained from previous studies<sup>11,15,16</sup> were submitted to the NCBI Web site, (<http://www.ncbi.nlm.nih.gov/>) using the BLAST program against the Transcriptome Shotgun Assembly (TSA) for *T. bicarinatum* (accession number GASM00000000.1) with the algorithm parameters adapted to short sequence tblastn searches. Putative complete ORFs were then obtained through the ExPASy Bioinformatics Resource Portal (<http://www.expasy.org/>) using a translate program. When the predicted ORFs were complete, the prepro-peptide sequences were deduced and used to search new, unknown peptide sequences with the BLAST program on the NCBI Web site. Generally, the ORFs were incomplete or shifted. The contigs containing partial or shifted ORFs were then reassembled using the blast2q and clustal Omega programs. The mass of mature peptide sequences obtained from these different approaches was systematically verified using the peptide mass program and compared to those obtained through mass spectrometry. Secondary structure predictions were conducted with the EMBOSS 6.3.1: Garnier program. Signal sequence and transmembrane domain were predicted with the phobius program available on <http://phobius.sbc.su.se/>. Phylogenetics trees were generated with MEGA software (megacc-7.0.26–1.x86\_64) with the Minimum Evolution (ME) or Maximum of Likelihood (ML) statistical methods, p-distance or Poisson model and edited with Seaview platform (1:4.4.2–1).

### 3. RESULTS

#### 3.1. Venom Mass Fingerprinting of *T. bicarinatum* Venom

First, a list of the most abundant peptides in the venom of *T. bicarinatum* was established through an initial LC–MS analysis on a LCQ-ion trap mass spectrometer. Prior to the determination of their sequences, all of the peptides detected were tentatively named according to the initials of the genus and species followed by their molecular weight (i.e., Tb-XXXX) as described by Johnson et al.<sup>17</sup> The fully sequenced peptides were subsequently renamed in accordance with the nomenclature developed for venom peptides.<sup>18</sup> The LC–MS analysis revealed 48 masses corresponding to peptides in the venom of *T. bicarinatum* (Figure 1 and Table 1). As expected for an ant venom,<sup>6</sup> all of the peptides were relatively small and exhibited molecular weights ranging from 939 to 4018 Da. These peptides eluted at retention times between 2.27 and 40.1 min which is equivalent to 6 and 35% acetonitrile in the mobile phase. The relative abundance of the 48 peptides is reported in Table 1. The ten most abundant peptides (i.e., Tb-2212 (bicarinatin), Tb-1571 (P17), Tb-2485, Tb-3325, Tb-1651, Tb-2638, Tb-4018, Tb-3882, Tb-1207, and Tb-1934) appeared to account for more than 67% of the venom peptide content. Although this peptide number was consistent with the previous investigation performed on the venom of *T. bicar-*



**Figure 1.** Positive mode total ion chromatogram of crude *T. bicarinatum* venom using an LCQ ESI mass spectrometer. The venom was separated onto a C<sub>18</sub> column. The mobile phase was 0.1% aqueous FA (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The peptides were eluted using a linear gradient from 0 to 50% B during 45 min, from 50 to 100% B during 10 min and then held for 5 min at a 200  $\mu\text{L min}^{-1}$  flow rate.

*inatum*<sup>14</sup> and on myrmicine ant venoms,<sup>6,19</sup> the HPLC coupled to highly sensitive mass spectrometry extended the numbers of peptides detected to hundreds and even to thousands peptides in the crude venoms of ants,<sup>7,20</sup> spiders<sup>21</sup> or cone snails.<sup>22</sup> Thus, we hypothesized that the venom of *T. bicarinatum* might be more complex than expected and we further analyzed its peptide content using a cutting-edge Orbitrap mass spectrometer. This resulted in a total number of 2853 peptides detected in the 1000–3000 Da mass range for 67% of the detected masses (data not shown). The molecular weight of the 48 previously detected peptides was adjusted accordingly in Table 1.

Then, we focused on the sequence identification of the venom peptides using a combination of direct and *de novo* sequencing. First, the venom was separated using a reversed-phase HPLC with a C<sub>18</sub> column for Edman degradation sequencing on purified peptides. Chromatographic separation allowed the effective purification of four peptides (i.e., Tb-2081, Tb-2671, Tb-3187 and Tb-4018). At the same time, crude venom was submitted to LC–MS/MS using an Orbitrap mass spectrometer for a *de novo* sequencing which yielded 1,778 sequence tags (data not shown). Most of these peptide tags were actually linked to a limited peptide diversity and corresponded to fragmented and truncated forms of mature peptides.

### 3.2. Identification of Venom Precursors, Peptide Toxin Classification and Nomenclature

The venomics methods are briefly summarized in Table S1. Overall, the venomics approach allowed us to assign 43 peptide sequences to the 48 peptide masses initially detected (Table 2). These peptide toxins are encoded by a total number of 29 precursors (including four which are incomplete). Moreover, the tblastn of the TSA of *T. bicarinatum* led to eight additional and full-length putative precursors encoding for eight putative mature peptides (Table S2) increasing the total number of peptides precursors to 37 units (Figure 2).

The alignment of the precursors sequences permitted us to cluster the venom peptide encoding gene transcripts (Figure 2). On the basis of the prepro-peptide sequence identity, the 37 venom peptide precursors were categorized into three different superfamilies (A–C). Subsequently, the examination of the mature region showed that the peptide toxins could be further separated into 21 groups or singletons based on amino acid sequence identity (Table 2 and Table S2). We assigned 20

**Table 1. Peptide Mass Fingerprint of *Tetramorium bicarinatum* Venom<sup>a</sup>**

retention time (min)	mass (Da)	relative abundance (%)	temporary name
2.27	<b>1555.8</b>	1.33	Tb-1556
2.88	1428.0	0.15	Tb-1428
7.39	2081.1	0.81	Tb-2081
11.97	2785.4	0.26	Tb-2785
12.04	2734.5	0.01	Tb-2734
12.35	2600.0	0.25	Tb-2600
12.75	2670.5	0.89	Tb-2671
<b>12.82</b>	<b>2467.4</b>	<b>1.01</b>	<b>Tb-2467</b>
13.28	1523.9	0.40	Tb-1524
<b>13.70</b>	<b>1934.0</b>	<b>2.04</b>	<b>Tb-1934</b>
15.49	1064.7	0.77	Tb-1065
15.91	3074.1	0.17	Tb-3074
<b>16.41</b>	<b>3186.6</b>	<b>1.32</b>	<b>Tb-3187</b>
<b>18.05</b>	<b>1052.6</b>	<b>1.81</b>	<b>Tb-1053</b>
18.18	1700.9	1.63	Tb-1701
18.98	1714.9	0.98	Tb-1715
<b>19.45</b>	<b>1295.7</b>	<b>1.95</b>	<b>Tb-1296</b>
<b>19.65</b>	<b>2423.2</b>	<b>1.18</b>	<b>Tb-2423</b>
19.93	1460.8	0.94	Tb-1461
<b>20.40</b>	<b>2604.4</b>	<b>1.28</b>	<b>Tb-2604</b>
20.55	2323.3	0.37	Tb-2323
20.68	2030.0	0.33	Tb-2030
20.82	1619.0	0.94	Tb-1619
21.09	2570.4	0.39	Tb-2570
<b>21.09</b>	<b>2773.5</b>	<b>1.32</b>	<b>Tb-2774</b>
21.85	2816.3	0.30	Tb-2816
21.85	2649.5	0.51	Tb-2650
<b>21.85</b>	<b>2852.6</b>	<b>1.55</b>	<b>Tb-2853</b>
22.13	3135.4	0.90	Tb-3135
22.46	1426.8	0.22	Tb-1427
22.74	1795.1	0.20	Tb-1795
23.90	1821.0	0.22	Tb-1821
<b>24.11</b>	<b>4018.1</b>	<b>4.99</b>	<b>Tb-4018</b>
26.13	939.3	1.68	Tb-0939
<b>26.27</b>	<b>1571.3</b>	<b>11.40</b>	<b>Tb-1571</b>
<b>27.20</b>	<b>2212.3</b>	<b>13.28</b>	<b>Tb-2212</b>
<b>27.52</b>	<b>2008.4</b>	<b>1.59</b>	<b>Tb-2008</b>
28.19	2004.2	0.27	Tb-2004
<b>29.09</b>	<b>3019.7</b>	<b>1.83</b>	<b>Tb-3020</b>
32.41	3324.8	7.73	Tb-3325
33.71	3871.3	1.43	Tb-3871
33.91	1650.9	7.40	Tb-1651
34.31	3882.2	2.96	Tb-3882
36.16	1075.7	0.30	Tb-1076
<b>37.37</b>	<b>2637.6</b>	<b>6.23</b>	<b>Tb-2638</b>
<b>38.30</b>	<b>1206.7</b>	<b>2.13</b>	<b>Tb-1207</b>
39.03	1408.7	1.03	Tb-1409
<b>40.10</b>	<b>2484.5</b>	<b>9.36</b>	<b>Tb-2485</b>

<sup>a</sup>List of peptide masses detected through LC–MS using an LCQ Advantage mass spectrometer. Bold indicates peptides with a relative abundance >1% of all the venom analyzed.

groups of peptide sequences with unknown biological activity and pharmacological targets a “U<sub>1</sub>–U<sub>20</sub>” prefix, while the membrane-active peptide called “bicarinalin”<sup>14,16</sup> was denoted with an “M” prefix. We chose to follow the rational nomenclature specifically developed for naming ant venom peptides that defined Myrmicitin (MYRTX) as the reference name suggested for the toxin peptides isolated from the ant

**Table 2.** Myrmicitoxin Sequences in the Venom of *Tetramorium bicarinatum* Venom

temporary name	proposed toxin name	relative abundance (%)	sequence	PTMs	net charge <sup>a</sup>	hydrophobic aa (%) <sup>b</sup>	pI <sup>c</sup>
Tb-2212	M-MYRTX-Tb1a (bicarinalin)	13.28	KIKIPWGKVKDVLVGGMKAV	NH <sub>2</sub>	5.0	50.00	10.18
Tb-1571	U <sub>1</sub> -MYRTX-Tb1a (P17)	11.40	LFKEILEKIKAKL	NH <sub>2</sub>	3.0	53.85	9.53
Tb-1053	U <sub>2</sub> -MYRTX-Tb1a	1.81	DPPPGFIGVR	NH <sub>2</sub>	1.0	30.00	5.84
Tb-2485	U <sub>3</sub> -MYRTX-Tb1a	9.36	VLPALPLLAGLMSLPLQHKLTN	NH <sub>2</sub>	2.1	56.52	8.73
Tb-1207	U <sub>3</sub> -MYRTX-Tb1a (cleaved)	2.13	VLPALPLLAGLM		0.0	75.00	5.49
Tb-1076	U <sub>3</sub> -MYRTX-Tb1a (cleaved)	0.30	VLPALPLLAGL		0.0	72.73	5.49
Tb-1428	U <sub>3</sub> -MYRTX-Tb1a (cleaved)	0.15	MSLPFLQHKLTN	NH <sub>2</sub>	2.1	41.67	8.52
Tb-1296	U <sub>3</sub> -MYRTX-Tb1a (cleaved)	1.95	SLPFLQHKLTN	NH <sub>2</sub>	2.1	36.36	8.49
Tb-3325	U <sub>3</sub> -MYRTX-Tb1b	7.73	IAPILALPLLLGGMMMSLPFLHHKLTGGKPHHE	NH <sub>2</sub>	2.4	45.16	8.62
Tb-1934	U <sub>3</sub> -MYRTX-Tb1b (cleaved)	2.04	SLPFLHHKLTGGKPHHE	NH <sub>2</sub>	2.4	23.23	8.40
Tb-1409	U <sub>3</sub> -MYRTX-Tb1b (cleaved)	1.03	IAPILALPLLLGGMM		0.0	71.43	5.52
Tb-2816	U <sub>3</sub> -MYRTX-Tb1b (cleaved)	0.30	ALPLLLGGMMMSLPFLHHKLTGGKPHHE	NH <sub>2</sub>	2.4	38.46	8.66
Tb-1524	U <sub>4</sub> -MYRTX-Tb1a	0.40	GCSFRRMRNLCG	NH <sub>2</sub> /1S-S	3.9	38.46	10.41
Tb-1556	U <sub>5</sub> -MYRTX-Tb1a	1.33	KRCEPDRARRGLC	NH <sub>2</sub> /1S-S	3.9	30.77	9.69
Tb-1701	U <sub>6</sub> -MYRTX-Tb1a	1.63	LWGKCPKIGGRRVMC	1S-S	3.9	46.57	10.10
Tb-1715	U <sub>6</sub> -MYRTX-Tb1b	0.98	LWGKCPKIGGRRIMC	1S-S	3.9	46.57	10.10
Tb-2030	U <sub>6</sub> -MYRTX-Tb1c	0.33	FRGCPKDMFKGRFIMC	1S-S	2.9	47.06	10.03
Tb-2081	U <sub>7</sub> -MYRTX-Tb1a	0.81	AINCRRYPRHPKCRGVSA	1S-S	5.0	38.89	10.86
Tb-1651	U <sub>8</sub> -MYRTX-Tb1a	7.40	GMLDRILGAVKGFMGS		1.0	50.00	8.75
Tb-2008	U <sub>9</sub> -MYRTX-Tb1a	1.59	GIVTKLIKKGVKLGLKMAL	NH <sub>2</sub>	6.0	52.63	10.60
Tb-3882	U <sub>10</sub> -MYRTX-Tb1a	2.96	GLGFLAKIMGKVGMRMIIKKLVPEAAKVAVDQLSQQQ		4.0	50.00	10.17
Tb-2323	U <sub>10</sub> -MYRTX-Tb1a (cleaved)	0.37	MIKKLVPEAAKVAVDQLSQQQ		1.0	47.62	8.25
Tb-4018	U <sub>11</sub> -MYRTX-Tb1a	4.99	GKEKEKLKQCFKDMTLAAIDYAKHKVEKHLFKCI	1S-S	4.1	44.12	9.30
Tb-0939	U <sub>12</sub> -MYRTX-Tb1a	1.68	LSPAVLASLA	NH <sub>2</sub>	1.0	70.00	10.00
Tb-2638	U <sub>13</sub> -MYRTX-Tb1a	6.23	RPPQIGIFDQIDKGMAAFMDLFK	NH <sub>2</sub>	1.0	47.83	6.04
Tb-2004	U <sub>13</sub> -MYRTX-Tb1a (cleaved)	0.27	RPPQIGIFDQIDKGMAAF		0.0	44.44	5.96
Tb-1065	U <sub>14</sub> -MYRTX-Tb1a	0.77	IPPNAVKSLQ	NH <sub>2</sub>	2.0	40.00	10.00
Tb-1619	U <sub>15</sub> -MYRTX-Tb1a	0.94	ILTADQLKAIKRH	NH <sub>2</sub>	3.1	50.00	9.99
Tb-1795	U <sub>15</sub> -MYRTX-Tb1b	0.20	VFLTPDQIKAMIKRH	NH <sub>2</sub>	3.1	46.07	9.99
Tb-3020	U <sub>16</sub> -MYRTX-Tb1a	1.83	GDEAGPKIGVFHDVNKAIEWLLQTK	NH <sub>2</sub>	2.1	37.04	8.39
Tb-2423	U <sub>16</sub> -MYRTX-Tb1a (cleaved)	1.18	GDEAGPKIGVFHDVNKAIEWL		-0.9	40.91	5.48
Tb-1821	U <sub>16</sub> -MYRTX-Tb1a (cleaved)	0.22	HDVNKAIEWLLQTK	NH <sub>2</sub>	2.1	40.00	8.51
Tb-2467	U <sub>17</sub> -MYRTX-Tb1a (nonglycosylated)	1.01	TIINAPNRCPGHHVVVKGRORIA	NH <sub>2</sub> /1S-S	5.0	43.48	10.79
Tb-2671	U <sub>17</sub> -MYRTX-Tb1a	0.89	TIINAPNRCPGHHVVVKGRORIA	NH <sub>2</sub> /O-glycosylation/1S-S	5.0	43.48	10.79
Tb-2570	U <sub>17</sub> -MYRTX-Tb1b (nonglycosylated)	0.39	TVIDVPIQCPSGTVKVGNKCRVIF	1S-S	2.9	45.83	8.88
Tb-2774	U <sub>17</sub> -MYRTX-Tb1b	1.32	TVIDVPIQCPSGTVKVGNKCRVIF	O-glycosylation/1S-S	2.9	45.83	8.88
Tb-2650	U <sub>17</sub> -MYRTX-Tb1c (nonglycosylated)	0.21	TIIDVPIQCPPGKVKGVRNCRVIF	1S-S	3.9	45.83	9.50
Tb-2853	U <sub>17</sub> -MYRTX-Tb1c	1.55	TIIDVPIQCPPGKVKGVRNCRVIF	O-glycosylation/1S-S	3.9	45.83	9.50
Tb-2604	U <sub>17</sub> -MYRTX-Tb1d	1.28	NIIRVPCRAYIEVNGVCREVFT	NH <sub>2</sub> /1S-S	1.9	52.17	8.06
Tb-3135	U <sub>17</sub> -MYRTX-Tb1e	0.90	NIIKAPLFPCPNGYIRDYKGDCREII	1S-S	0.9	44.44	6.20
Tb-3187	U <sub>17</sub> -MYRTX-Tb1f	1.32	NIIRVPEFQCPNGYRKDANGKCREVFH	NH <sub>2</sub> /1S-S	3.0	37.04	8.86
Tb-3074	U <sub>17</sub> -MYRTX-Tb1f (cleaved)	0.17	IIRVPEFQCPNGYRKDANGKCREVFH	NH <sub>2</sub> /1S-S	3.0	38.46	8.86
Tb-2785	U <sub>17</sub> -MYRTX-Tb1g	0.26	HVIDTRSRLCPEGSRSTTGECKTV	1S-S	2.0	24.00	8.96
Tb-2734	U <sub>17</sub> -MYRTX-Tb1h	0.01	HIIRVPCRAYKEIRGRCRKILT	NH <sub>2</sub> /1S-S	7.0	43.48	10.90

**Table 2.** continued

<sup>a</sup>According to PepCalc (<https://pepcalc.com/>). <sup>b</sup>According to Peptide Property Calculator V3.1 (<https://www.biosyn.com/peptidepropertycalculator/peptidepropertycalculator.aspx>). <sup>c</sup>According to expasy ([https://web.expasy.org/compute\\_pi/](https://web.expasy.org/compute_pi/)).

## **Superfamily-A**

U5-MYRTX-Tbla MQ-LSHLLAFAFVIVWV-TIYA-PQQADAMADA-KRCEPDRARRGLCG-  
 U4-MYRTX-Tbla MQ-PQSLLAFAFVIVWV-TIYA-PQAEEAKAGADADAXHADKA-GCSQFRRMRNLLCGGK  
 U7-MYRTX-Tbla MQ-LSHLLAFAFVIVWV-TIHT-PQOQADAMADADA-AHNCRKYRPHFKRCRVSA--  
 U18-MYRTX-Tbla MQ-LSHLLAFAFVIVWV-TIYA-PQOQADAMADADA-D-DVNECWEVTPYHPD-CRGVSILPWRWTKICYRC-  
 M-MYRTX-Tbla MK-LSPFSLAVLAIYLV-A-MLTY-PHAEEAKAWADADADA-KADADAYA-VADALADAVAKHPIWPGVKDFEVGGMKAVGK  
 U2-MYRTX-Tbla --EAAEAMADAMADAMADAMADAMADAMAEAAADPPGFIVGRG-  
 U1-MYRTX-Tbla MK-LSPFSLALATIFVVA-TIYA-PQMLARASSDADADAASASDADADAEASAS-LFKIELEKIKARPGKK-  
 U9-MYRTX-Tbla MK-LSPFSLALAIYVVA-TIYA-PQOKAKASADADADAASASNALAEASA-GHVTKLIIKGKVGLKLMALGKK-  
 U8-MYRTX-Tbla MK-LSPFSLAFIVVVA-TIYA-PQEAESASADADADAASASDALAKASA-GMDRLIGAVKGMFGMS-  
 U11-MYRTX-Tbla --LA-MAGCDPAAVDAKQDAAKRMAYAAYIA-GKEKEKILQOCFKDMTAAIDIYAHKVEKHLFCI-  
 U10-MYRTX-Tbla MR-VSY-SITLTIVVVA-HI-PETEAKAWADADA-GLFLAKIIMKGKVGMRIKKLVEPEAKVAVDQLSQQQ-  
 U6-MYRTX-Tbla --LA-VVYVTLA-TIVPTVYNAEADADA-LWGKCPKIGGR-R-VMC-  
 U6-MYRTX-Tblb --LA-VVYVTLA-TIVPTVYNAEADADA-LWGKCPKIGGR-R-IC-  
 U6-MYRTX-Tblc MLQALCFLS-TLTLVTL-TIHPMAAANAEPE-FRGPCKPDIMFKGRFT-  
 U3-MYRTX-Tbla MKVYKEPFIA-VIIVVGLSGSTWASPIANAKAEDAAEAEVAKVAKA-VIP-ALBLPLACMSDPFLQHKLNTG-  
 U3-MYRTX-Tblb MKVYKEPFIA-VIIVVGLSGSTWASPLAKAKAEDAAEAEAA-IAPILALPLLGGMMSDPFLHHKLTLGKPKPHHEG-  
 U3-MYRTX-Tblc MKVYKEPFIA-VIIVVGLSGSTWASPLAKAKAEDAAEAEAA-IAPIVALLLSGCFSPFLHHKLNTGMHHHEG-  
 Pilosulin1 MKLSCSSLTL-TIIPVYVLA-VH-A-PNVEAKDLASPESESVGADPFGED-DAVGEADPNAQDGSVFGRALIIGRVIPKVKAAVLGFVKVAKVLPKVMKEAIPMAVEAMAKSQEEQQPQ-  
 Pilosulin3A MKLSCSSLTL-A-IIIVFVLA-VH-A-PNVEAKLALAPESVDAVPEADPFGED-P-1DWKKWDVKKVSKK-TCKVMLKA--CKFLG-  
 Pilosulin4 MKLSCSSLTL-A-IIIVFVLA-VH-A-PNVEAKLALAPESVDAVPEADPFGED-P-FDTIKLNLKKLTKA-TCKVISKGASMCKVLFDKKKQE-  
 Pilosulin3 MKLSCSSLTL-A-IIIVFVLA-VH-A-PNVEAKLALAPESVDAVPEADPFGED-P-AN-1GLVSRGTCVLRVTCVKVVLQG-  
 Pilosulin5 MKLSCSSLAL-AIILIA-VH-SPNEEVKALADPEADPFGEDDAFAEA-NAVVKGMKKAIEKEDVIEKGYDKLAA--KLKKVIIQQLWE-

## Superfamily-B

U12-MYRTX-Tbl1 MKVKKLITIFAMIAALMVTAVAV-----GKAKSPVLA  
 U14-MYRTX-Tbl1 MKVKKLITIFTMMALMVTAVAV-----GEPFPNVWKSQCG  
 U15-MYRTX-Tbl1 MKVKKLITIFAMIAALMVTAVAV-----GEAVETTPDOKKAIKRHG  
 U15-MYRTX-Tbl1 MKVKKLITIFAMIAALMVTAVAV-----GEAVETTPDOKKAIKRHG  
 U19-MYRTX-Tbl1 MKSNLLITIFAMIAALMMAA-----GEAARSRLKIGRMER  
 U20-MYRTX-Tbl1 MKSBNLLITIFAMIAALMMAA-----GEAARVSDCMTTSK-LLKLLPV  
 U16-MYRTX-Tbl1 MK---LLYIPFLAVAVIAVPGIMGAGDEADPKAKGVPHDVNPAAEELLQQTKG  
 U13-MYRTX-Tbl1 MK---LLYIPFLAVAVIAVPGIMGAEZAEGRPDKGIFDQIDGAAAFMDLFKG

## **Superfamily-C**

U17-MYRTX-Tbla	M <b>E</b> KNRNT <b>N</b> IFS <b>V</b> YL <b>M</b> IT <b>F</b> LL <b>I</b> S <b>F</b> IP <b>T</b> TM <b>V</b> MS <b>D</b> G <b>E</b> A <b>T</b> I <b>N</b> AP <b>R</b> N <b>R</b> ---	<b>C</b> PP <b>G</b> H <b>-V</b> V <b>K</b> GR <b>C</b> <b>I</b> <b>A</b> G <b>-</b>
U17-MYRTX-Tblb	M <b>E</b> KNRNT <b>N</b> IFS <b>V</b> YL <b>M</b> IT <b>F</b> LL <b>I</b> S <b>F</b> IP <b>T</b> TM <b>V</b> MS <b>D</b> G <b>E</b> A <b>T</b> I <b>N</b> AP <b>R</b> N <b>R</b> ---	<b>E</b> SG <b>T</b> - <b>V</b> K <b>G</b> NC <b>V</b> <b>I</b> <b>S</b>
U17-MYRTX-Tblc	M <b>E</b> KHRNT <b>N</b> IFS <b>V</b> YL <b>M</b> IT <b>F</b> LL <b>I</b> S <b>F</b> IP <b>T</b> TM <b>V</b> MS <b>D</b> G <b>E</b> A <b>T</b> I <b>N</b> AP <b>R</b> N <b>R</b> ---	<b>E</b> PG <b>K</b> - <b>V</b> K <b>G</b> NC <b>V</b> <b>I</b> <b>S</b>
U17-MYRTX-Tbld	M <b>E</b> KNRNT <b>T</b> TF <b>S</b> VL <b>T</b> L <b>I</b> F <b>L</b> IS <b>T</b> IP <b>T</b> TM <b>V</b> MT <b>-</b> <b>E</b> AN <b>I</b> RV <b>P</b>	<b>R</b> AG <b>C</b> <b>I</b> <b>E</b> W <b>N</b> VC <b>E</b> EV <b>T</b>
U17-MYRTX-Tble	M <b>E</b> KNRNT <b>T</b> TF <b>S</b> VL <b>T</b> L <b>I</b> F <b>L</b> IS <b>T</b> IP <b>T</b> TM <b>V</b> MT <b>-</b> <b>E</b> SN <b>I</b> K <b>A</b> LP <b>L</b> F <b>P</b>	<b>C</b> EN <b>G</b> I <b>R</b> D <b>Y</b> K <b>D</b> <b>C</b> <b>E</b> <b>I</b> <b>L</b> E
U17-MYRTX-Tblk	M <b>E</b> KNRNT <b>T</b> TF <b>S</b> VL <b>T</b> L <b>I</b> F <b>L</b> IS <b>T</b> IP <b>T</b> TM <b>V</b> MT <b>-</b> <b>E</b> SN <b>I</b> R <b>V</b> EE <b>P</b> <b>F</b>	<b>C</b> EN <b>G</b> I <b>R</b> D <b>K</b> D <b>A</b> <b>C</b> <b>E</b> <b>P</b> <b>H</b>
U17-MYRTX-Tblg	M <b>E</b> KSR <b>S</b> RC <b>G</b> FF <b>S</b> YL <b>M</b> TL <b>L</b> IS <b>T</b> IP <b>T</b> TM <b>V</b> MT <b>-</b> <b>E</b> HV <b>I</b> <b>T</b> RS <b>R</b> L	<b>C</b> EG <b>S</b> R <b>S</b> RT <b>T</b> <b>C</b> <b>E</b> <b>T</b> <b>V</b>
U17-MYRTX-Tblh	M <b>E</b> K-R <b>V</b> G <b>T</b> <b>S</b> YL <b>M</b> TL <b>L</b> IS <b>T</b> IP <b>T</b> TM <b>V</b> MT <b>-</b> <b>E</b> H <b>I</b> RV <b>P</b>	<b>R</b> AG <b>C</b> <b>I</b> <b>-</b> <b>K</b> E <b>R</b> <b>G</b> <b>I</b> <b>L</b> <b>T</b>
U17-MYRTX-Tblj	M <b>E</b> KSP <b>T</b> TF <b>S</b> YL <b>M</b> TL <b>L</b> IS <b>T</b> IP <b>T</b> TM <b>V</b> MT <b>-</b> <b>E</b> RS <b>V</b> <b>R</b> AR <b>E</b> <b>R</b>	<b>C</b> PS <b>S</b> <b>Q</b> <b>M</b> D <b>G</b> <b>S</b> <b>C</b> <b>K</b> <b>G</b> <b>R</b> <b>E</b>
U17-MYRTX-Tblk	M <b>E</b> KNRNT <b>N</b> IFS <b>V</b> YL <b>M</b> IT <b>F</b> LL <b>I</b> S <b>F</b> IP <b>T</b> TM <b>V</b> MS <b>D</b> <b>S</b> Y <b>I</b> RV <b>T</b> FP <b>P</b> P <b>T</b> <b>O</b> PP <b>G</b> E <b>-</b> <b>T</b> W <b>M</b> <b>Y</b> <b>K</b> R <b>G</b> <b>C</b> <b>H</b> <b>V</b>	<b>C</b> EN <b>G</b> I <b>R</b> D <b>Y</b> <b>E</b> <b>C</b> <b>D</b> <b>E</b> <b>I</b> <b>F</b> <b>P</b>
U17-MYRTX-Tbll	M <b>E</b> KSR <b>S</b> ST <b>F</b> S <b>I</b> YL <b>M</b> TL <b>F</b> LV <b>S</b> T <b>P</b> TM <b>V</b> MT <b>-</b> <b>E</b> SD <b>I</b> BA <b>P</b> PF <b>C</b> <b>E</b> <b>D</b> <b>L</b> <b>V</b> PL <b>R</b> S <b>-</b> <b>K</b> <b>C</b> <b>E</b> <b>S</b> R <b>M</b> <b>S</b> <b>I</b> <b>C</b> <b>Q</b> <b>R</b> <b>K</b> <b>V</b> <b>S</b> <b>K</b> <b>R</b>	

**Figure 2.** Multiple alignment of *T. bicarinatum* venom peptide precursors in superfamilies -A, -B and -C. Alignments were generated with the muscle program in Seaview 1:4.5.4.8-2 and edited using BOXSHADE 3.3.1-9. Gaps were introduced to optimize the alignments. Pilosulin precursors from *Myrmecia pilosula*, which have homology sequences with MYRTXs, have been included. Identical residues are highlighted in magenta. Similar residues in the peptide sequences are highlighted in blue while conserved residues are shown in cyan. Black triangles indicate the cleavage site releasing mature peptides. \*: For prepro-U<sub>13</sub>-MYRTX-Tb1a, the cleavage site producing mature peptides occurs after the GEAEAEG motif. The predicted signal sequences are underlined.

subfamily Myrmicinae.<sup>5</sup> Then, the toxin name was followed with the initial of the genus and species (Tb) and an alpha-numeric code to distinguish different isotoxins in the same group (i.e., 1a, 1b, 1c).

Finally, we calculated the theoretical molecular weight of each putative mature peptide taking into account the post-translational modifications that were detected through mass spectrometry or deduced from their precursor sequences ([Table S1](#) and [Table S2](#)). The presence of peptides in the crude venom was validated by matching the masses observed in the mass spectrometry with those predicted from both transcriptomic and proteomic sequencing data. The expression of genes encoding the precursors in the venom glands was also confirmed by RT-PCR for 29 of the 37 precursors ([Table S1](#)). Below, we describe the Myrmicitoxins identified in the present study, their precursor superfamily, their precursor processing, their post-translational modifications, and their homologies with other hymenoptera toxins.

### 3.2.1. Superfamily-A: Pilosulin-like Peptide.

superfamily was made up of 17 MYRTX precursors (preproto-M-Tb1a, -U<sub>1</sub>-Tb1a, -U<sub>2</sub>-Tb1a, -U<sub>3</sub>-Tb1a, -U<sub>3</sub>-Tb1b, -U<sub>3</sub>-Tb1c, -U<sub>4</sub>-Tb1a, -U<sub>5</sub>-Tb1a, -U<sub>6</sub>-Tb1a, -U<sub>6</sub>-Tb1b, -U<sub>6</sub>-Tb1c, -U<sub>7</sub>-Tb1a, -U<sub>8</sub>-Tb1a, -U<sub>9</sub>-Tb1a, -U<sub>10</sub>-Tb1a, -U<sub>11</sub>-Tb1a and -U<sub>18</sub>-Tb1a), including four partially determined precursors (preproto-

$U_2$ -Tb1a,  $-U_6$ -Tb1a,  $-U_6$ -Tb1b and  $-U_{11}$ -Tb1a) (Figure 2). The consensus sequence of the prepro-region of the 13 full-length precursors was defined as MKVLSXLXAXIFVMX-XXXWAXPXXXAKAXADADADA. The prepro-regions of these precursors shared about 50% identity with the consensus sequence of the superfamily-A even though the prepro- $U_3$ -Tb1a,  $-U_3$ -Tb1b,  $-U_3$ -Tb1c and  $-U_6$ -Tb1c precursors were more dissimilar as they shared less than 38% identity with the consensus sequence. Consequently, we subdivided the superfamily-A into three families as follows. The A1 family is composed of prepro-M-Tb1a,  $-U_1$ -Tb1a,  $-U_2$ -Tb1a,  $-U_4$ -Tb1a,  $-U_5$ -Tb1a,  $-U_6$ -Tb1a,  $-U_6$ -Tb1b,  $-U_7$ -Tb1a,  $-U_8$ -Tb1a,  $-U_9$ -Tb1a,  $-U_{10}$ -Tb1a,  $-U_{11}$ -Tb1a, and  $-U_{18}$ -Tb1a. The A2 family includes the prepro- $U_3$ -Tb1a,  $-U_3$ -Tb1b, and  $-U_3$ -Tb1c and the A3 family is represented by the unique full-length precursor prepro- $U_6$ -Tb1c (Table S3).

This superfamily was closely related to the pilosulin precursors described by Inagaki<sup>23</sup> which shared 35% sequence identity with the consensus prepro-sequence of superfamily-A and 44% sequence identity with the consensus sequence of A1 family (Figure 2).

Most of the superfamily-A pro-peptides bore the conserved motif ADADA with an additional C-terminal extension (between 0 to 22 aa) before the mature sequence. The

Secapin Family	10	20	30	% ID	Species
U17-MYRTX-Tb1a	T <small>I</small> I <small>N</small> A <small>P</small> ---	N <small>R</small> C <small>P</small> P <small>G</small> H-V <small>V</small> V <small>K</small> G <small>R</small> C <small>R</small> I <small>A</small> *		-	<i>Tetramorium bicarinatum</i>
U17-MYRTX-Tb1b	T <small>V</small> I <small>D</small> V <small>P</small> ---	I <small>Q</small> C <small>P</small> S <small>G</small> T-V <small>K</small> V <small>G</small> N <small>K</small> C <small>R</small> V <small>I</small> F		43.5	<i>Tetramorium bicarinatum</i>
U17-MYRTX-Tb1c	T <small>I</small> I <small>D</small> V <small>P</small> ---	I <small>Q</small> C <small>P</small> P <small>G</small> K-V <small>K</small> V <small>G</small> N <small>R</small> C <small>R</small> V <small>I</small> F		56.5	<i>Tetramorium bicarinatum</i>
U17-MYRTX-Tb1d	N <small>I</small> I <small>R</small> V <small>P</small> ---	C <small>R</small> A <small>G</small> Y-I <small>E</small> V <small>N</small> G <small>V</small> C <small>R</small> E <small>V</small> F <small>T</small> *		39.1	<i>Tetramorium bicarinatum</i>
U17-MYRTX-Tb1e	N <small>I</small> I <small>K</small> A <small>P</small> L---	F <small>P</small> C <small>P</small> N <small>G</small> Y-I <small>R</small> D <small>Y</small> K <small>G</small> D <small>C</small> R <small>E</small> I <small>I</small> E*		44.0	<i>Tetramorium bicarinatum</i>
U17-MYRTX-Tb1f	N <small>I</small> I <small>R</small> V <small>P</small> E---	F <small>O</small> C <small>P</small> N <small>G</small> Y-R <small>K</small> D <small>A</small> N <small>G</small> K <small>C</small> R <small>E</small> V <small>F</small> H*		36.0	<i>Tetramorium bicarinatum</i>
U17-MYRTX-Tb1g	H <small>V</small> I <small>D</small> T---	R <small>S</small> R <small>L</small> C <small>P</small> E <small>G</small> S <small>R</small> R <small>S</small> T <small>T</small> G <small>E</small> C <small>K</small> T <small>V</small>		23.1	<i>Tetramorium bicarinatum</i>
U17-MYRTX-Tb1h	H <small>I</small> I <small>R</small> V <small>P</small> ---	C <small>R</small> A <small>G</small> Y-K <small>E</small> I <small>R</small> G <small>R</small> C <small>R</small> K <small>I</small> L <small>T</small> *		39.1	<i>Tetramorium bicarinatum</i>
U17-MYRTX-Tb1i	R <small>V</small> I <small>D</small> A---	R <small>E</small> R <small>C</small> P <small>S</small> G <small>Y</small> Q <small>M</small> D <small>G</small> S <small>C</small> K <small>C</small> R <small>K</small> I <small>F</small> G <small>R</small>		36.0	<i>Tetramorium bicarinatum</i>
U17-MYRTX-Tb1j	Y <small>I</small> I <small>R</small> V <small>P</small> T-F <small>P</small> P <small>P</small> T-C <small>P</small> P <small>G</small> E-T <small>M</small> V <small>G</small> K <small>R</small> C <small>R</small> H <small>V</small> Y			40.7	<i>Tetramorium bicarinatum</i>
U17-MYRTX-Tb1k	Y <small>I</small> I <small>E</small> A <small>P</small> P---	F <small>P</small> C <small>P</small> N <small>G</small> Y-M <small>R</small> D <small>Y</small> E <small>G</small> D <small>C</small> R <small>E</small> I <small>F</small> E*		40.0	<i>Tetramorium bicarinatum</i>
U17-MYRTX-Tb1l	D <small>I</small> I <small>D</small> V <small>P</small> L-R <small>V</small> S <small>K</small> C <small>P</small> E <small>G</small> S <small>R</small> M <small>S</small> T-I <small>G</small> O <small>C</small> R <small>K</small> V <small>K</small> S <small>R</small>			33.3	<i>Tetramorium bicarinatum</i>
secapin	Y <small>I</small> I <small>D</small> V <small>P</small> ---	P <small>R</small> C <small>P</small> P <small>G</small> S-K <small>F</small> I <small>K</small> N <small>R</small> C <small>R</small> V <small>I</small> V <small>P</small>		52.2	<i>Apis mellifera</i>
secapin	Y <small>I</small> I <small>N</small> V <small>P</small> ---	P <small>R</small> C <small>P</small> P <small>G</small> S-K <small>F</small> I <small>K</small> N <small>R</small> C <small>R</small> V <small>I</small> V <small>P</small>		52.2	<i>Polistes hebraeus</i>
secapin	Y <small>I</small> I <small>D</small> V <small>P</small> ---	P <small>R</small> C <small>P</small> P <small>G</small> S-K <small>F</small> V <small>H</small> K <small>R</small> C <small>R</small> V <small>I</small> V <small>P</small>		56.5	<i>Vespa maculifrons</i>

Poneratoxin Family	10	20	% ID	Species
U3-MYRTX-Tb1a	V <small>L</small> P <small>A</small> L <small>P</small> L <small>L</small> A <small>G</small> L <small>M</small> S <small>L</small> P <small>F</small> L <small>Q</small> -H <small>K</small> L <small>T</small> N*		-	<i>Tetramorium bicarinatum</i>
δ-PPONTX-Pc1c	F <small>L</small> P <small>I</small> I <small>L</small> I <small>L</small> G <small>S</small> L <small>M</small> T <small>P</small> P <small>V</small> I <small>Q</small> A <small>I</small> H <small>D</small> E <small>Q</small> R*		36.0	<i>Paraponera clavata</i>
δ-PPONTX-Pc1b	F <small>L</small> P <small>I</small> I <small>L</small> I <small>L</small> G <small>S</small> L <small>M</small> T <small>P</small> P <small>V</small> I <small>Q</small> A <small>I</small> H <small>D</small> A <small>Q</small> R		36.0	<i>Paraponera clavata</i>
δ-PPONTX-Pc1d	F <small>L</small> P <small>I</small> I <small>L</small> I <small>L</small> G <small>S</small> L <small>M</small> T <small>P</small> P <small>V</small> I <small>Q</small> A <small>I</small> H <small>N</small> V <small>Q</small> R*		36.0	<i>Paraponera clavata</i>
δ-PPONTX-Pc1a	F <small>L</small> P <small>I</small> I <small>L</small> I <small>L</small> G <small>S</small> L <small>M</small> T <small>P</small> P <small>V</small> I <small>Q</small> A <small>I</small> H <small>D</small> A <small>Q</small> R		36.0	<i>Paraponera clavata</i>

Ponericin Family	10	20	30	% ID	Species
U10-MYRTX-Tb1a	G <small>I</small> G <small>F</small> L <small>A</small> K <small>I</small> M <small>G</small> K <small>V</small> G <small>M</small> R <small>M</small> I <small>K</small> K <small>L</small> V <small>P</small> E <small>A</small> A <small>K</small> V <small>A</small> V <small>D</small> Q <small>L</small> S <small>Q</small> Q			-	<i>Tetramorium bicarinatum</i>
M-PONTX-Ng3g	G-L <small>V</small> D <small>V</small> L <small>G</small> K <small>V</small> G <small>G</small> -L <small>I</small> K <small>K</small> L <small>L</small> P <small>G</small>			55.0	<i>Neoponera goeldii</i>
M-PONTX-Ng3f	G-L <small>V</small> D <small>V</small> L <small>G</small> K <small>V</small> G <small>G</small> -L <small>I</small> K <small>K</small> L <small>L</small> P <small>G</small> *			52.4	<i>Neoponera goeldii</i>

Uncategorized Family	10	% ID	Species
U12-MYRTX-Tb1a	L <small>S</small> P <small>A</small> V <small>I</small> A <small>S</small> L <small>A</small> *	-	<i>Tetramorium bicarinatum</i>
OdVP4	L <small>D</small> P <small>K</small> V <small>V</small> Q <small>S</small> L <small>L</small> *	50.0	<i>Orancistrocerus drewseni</i>
EpVP4b	L <small>S</small> P <small>A</small> A <small>M</small> A <small>S</small> L <small>A</small> *	80.0	<i>Eumenes pomiformis</i>
EpVP4a	L <small>S</small> P <small>A</small> V <small>M</small> A <small>S</small> L <small>A</small> *	90.0	<i>Eumenes pomiformis</i>
EpVP3s	I <small>N</small> P <small>K</small> S <small>V</small> Q <small>S</small> L <small>L</small> *	30.0	<i>Eumenes pomiformis</i>
EpVP3	A <small>I</small> N <small>P</small> K <small>S</small> V <small>Q</small> S <small>L</small> L*	30.0	<i>Eumenes pomiformis</i>

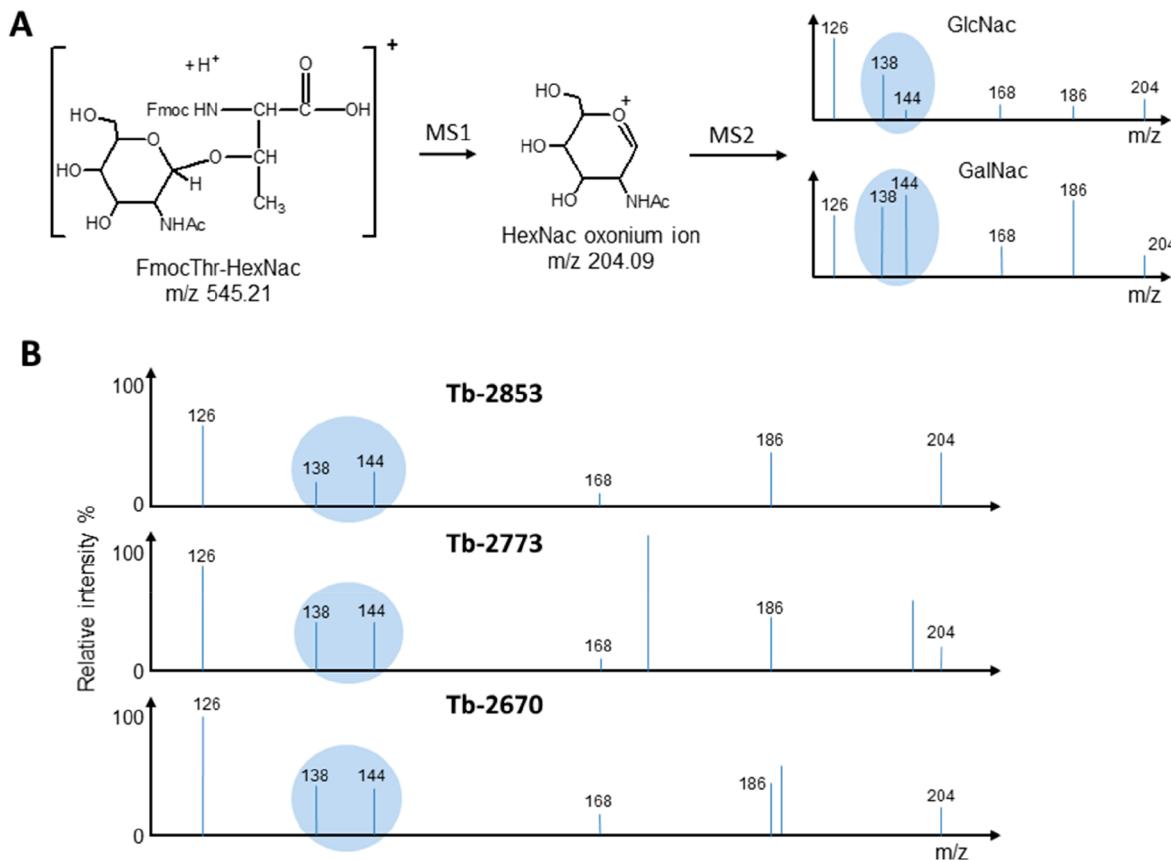
**Figure 3.** Multiple sequence alignment of mature myrmicotoxins in *Tetramorium bicarinatum* venom with homologous venom peptides from hymenoptera. Resulting alignments using the ClustalX program were edited with BOXSHADE 3.3.1–9. Gaps were introduced to optimize the alignments. Identical residues are highlighted in magenta. Similar residues in the peptide sequences are highlighted in blue while conserved residues are shown in cyan. Percentage identity (% ID) is relative to the first peptide of each family. Red stars represent C-terminal amidation and the threonine residues inside red box are O-glycosylated.

prepro-peptide sequences were submitted to the Phobius program to predict the transmembrane and signal sequence domains. For all of the precursors of superfamily-A, the N-terminal (from 1 to 24–29 position) was predicted to be a signal sequence while the remaining part that included both pro-region and mature peptides was predicted to be a noncytosolic domain. The cleavage site between the pro-peptide and the mature region was a conserved XAXA motif except for U<sub>3</sub>-Tb1c and U<sub>6</sub>-Tb1c (Figure 2). Nine MYRTX precursors possessed the GKK motif or G as an extra C-terminal signal for amidation.

The mature peptides belonging to the superfamily-A of precursors had quite divergent primary sequences with no strictly conserved features in terms of amino acid composition, molecular size or cysteine residues (Table 2). Interestingly, two peptides described here had sequence homologies with other ant venom peptides. The peptide U<sub>3</sub>-MYRTX-Tb1a possessed 36% sequence identity with the poneratoxin ( $\delta$ -PPONTX-Pc1a) a 25-residue linear peptide from *Paraponera clavata* while the peptide U<sub>10</sub>-MYRTX-Tb1a had 55 and 52% sequence identity with ponerincin G6 and G7 (M-PONTX-Ng3f and M-PONTX-Ng3g) isolated from the venom of *Neoponera goeldii* (Figure 3). Except for U<sub>3</sub>-MYRTX-Tb1a and U<sub>10</sub>-MYRTX-Tb1a, the other peptides described in the superfamily-A of precursors had no similarities with previously described ant venom peptides. Interestingly, despite being

assigned to pilosulin precursors, superfamily-A's mature peptides were very divergent. The sequence analysis of mature peptides from the superfamily-A also revealed that several short peptides were fragments of larger peptides and belonged to the same toxin precursor. The U<sub>3</sub>-MYRTX group of peptides is produced by three homologous peptide precursors that yield the full-length mature peptides (U<sub>3</sub>-Tb1a, U<sub>3</sub>-Tb1b and U<sub>3</sub>-Tb1c) (Table 2 and Table S3), plus additional truncated versions of these peptides (Tb-1207, Tb-1076, Tb-1428, Tb-1296, Tb-1934, Tb-1409 and Tb-2816) (Table 2). As for U<sub>3</sub>-MYRTXs, we discovered that Tb-2323 was also a fragment of the peptide U<sub>10</sub>-Tb1a (Table 2). Therefore, the maturation of 17 precursors of the superfamily-A generated 22 peptides that were confirmed by mass spectrometry as present in the venom (Table 2). An additional putative mature peptide with two disulfide bonds was predicted to be generated by the U<sub>18</sub>-MYRTX-Tb1a precursor even though its presence in the crude venom was not confirmed through mass spectrometry (Table S2).

**3.2.2. Superfamily-B: New Ant Venom Precursors.** The superfamily-B, which included eight precursors (Figure 2), was subdivided into the B1 (prepro-U<sub>12</sub>-Tb1a, -U<sub>14</sub>-Tb1a, -U<sub>15</sub>-Tb1a, -U<sub>15</sub>-Tb1b, -U<sub>19</sub>-Tb1a and -U<sub>20</sub>-Tb1a) and B2 (prepro-U<sub>16</sub>-Tb1a and -U<sub>13</sub>-Tb1a) families (Table S3). The precursors had a consensus prepro-sequence defined as MK(I/L/S)I(K/N/Y)LITIFAMIATLMXTXXGEA and shared about



**Figure 4.** Fragmentation patterns of HexNAc-threonine and in situ-generated oxonium ions (A). Saccharide oxonium ion spectra of native U17-Tb1c (Tb-2853), U17-Tb1b (Tb-2774) and U17-Tb1a (Tb-2671) (B).

63% identity with this consensus sequence. The consensus sequence showed 13% identity with the consensus prepro-sequence of the superfamily-A and had no significant homology to any previously known ant venom peptide sequence. Consequently, the superfamily-B should be considered as a new ant venom precursor superfamily.

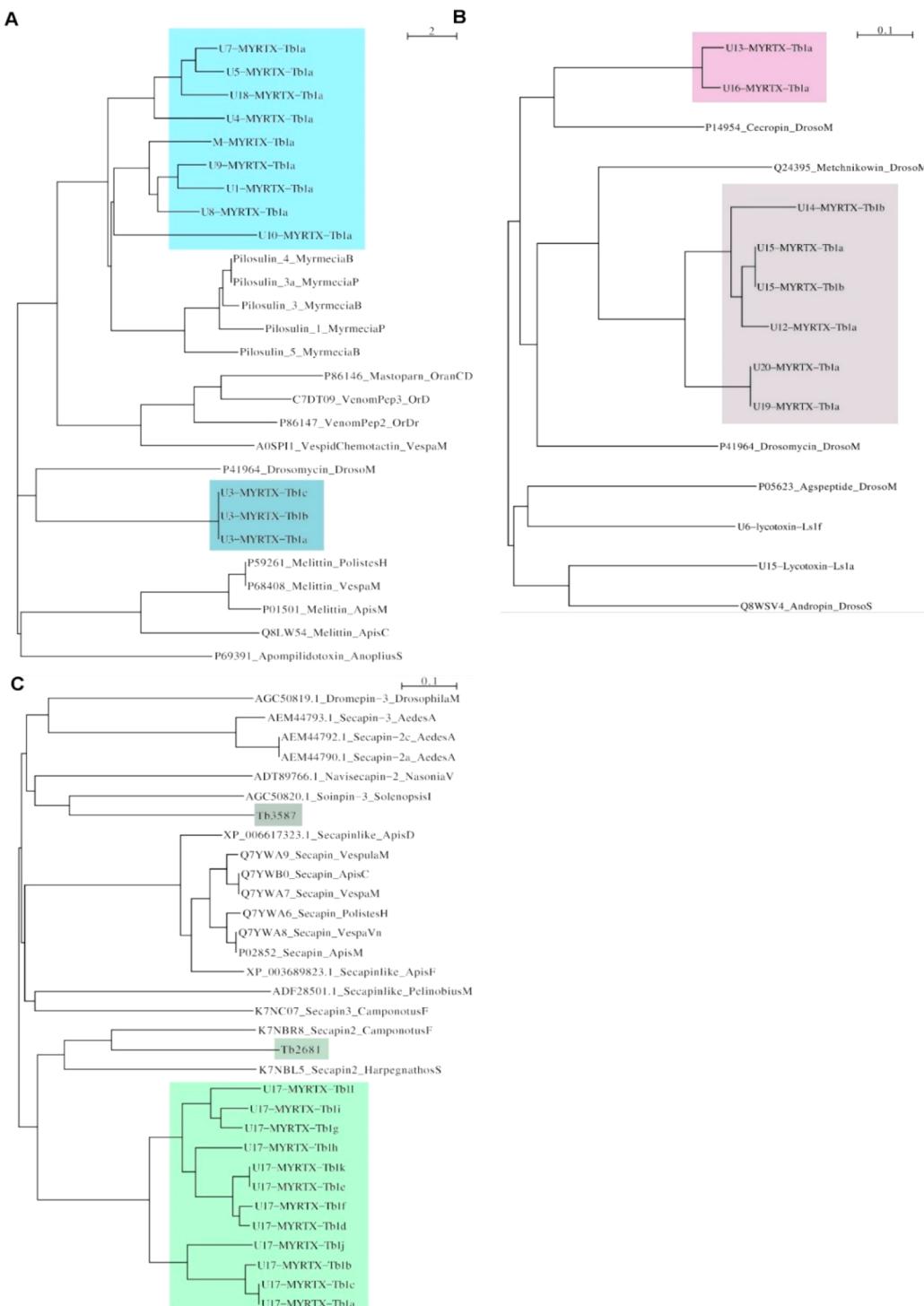
The expression of prepro-U<sub>19</sub>-Tb1a and -U<sub>20</sub>-Tb1a in the venom glands was confirmed through RT-PCR even though the theoretical masses of their mature peptides were not detected in the crude venom using mass spectrometry. The first 25 amino acids at the N-terminus of these precursors were predicted as a signal sequence by the Phobius program while the remaining C-terminal part corresponding to mature peptide was predicted as a noncytosolic domain. The mature peptides were cleaved after a XGEA, NGEP, EAEG or NGKA motif to produce peptides ranging from 10 to 27 amino acids (Figure 2). Additional cleavages of the two mature peptides U<sub>13</sub>-Tb1a and U<sub>16</sub>-Tb1a resulted in their corresponding short versions Tb-2004, and Tb-2423 and Tb-1821, respectively (Table 2).

There was no significant similarity between the precursors from superfamily-B and other venom precursors. However, despite a very distinct prepro-region (only 28% identity), the mature peptide U<sub>12</sub>-MYRTX-Tb1a and the peptide VP4a, isolated from the solitary wasp *Eumenes pomiformis* (UniProt accession number: D1MEJO), are very similar to 90% of sequence identity (Figure 3).

**3.2.3. Superfamily-C: Secapin-like Peptides.** The 12 precursors of the superfamily-C (Figure 2) yielded MYRTXs that shared significant sequence similarity with the secapins, a

group of multifunctional peptides found in the venom of numerous hymenoptera (Figure 3). However, the prepro-region of these precursors showed no clear sequence similarity with other insect secapin precursors (12% identity on average between consensus prepro-sequence and other insect secapin prepro-sequences). Both the prepro-region and mature sequences of superfamily-C precursors were highly conserved (Figure 2). The consensus prepro-sequence was defined as MEKNRTXTFSXYLXITXXLISTFITMVIT(DG)EX and the consensus cleavage site as “TMV(IT/MS)(DG)E(A/S/G)”. The percentage of sequence identity was about 75% between prepro-sequences and consensus sequence. The presence in the crude venom of the four mature peptides (i.e., U<sub>17</sub>-Tb1i, U<sub>17</sub>-Tb1j, U<sub>17</sub>-Tb1k and U<sub>17</sub>-Tb1l) was not confirmed through proteomic analysis (Table S2). The 12 mature peptides contained from 23 to 30 amino acids and all contained one disulfide bond. Six peptides were amidated at the C-terminus with an extra glycine residue in the precursor. An alignment of the sequence of the 12 secapin-like peptides in the venom of *T. bicarinatum* revealed several conserved features including the two cysteine residues, a pair of isoleucines at the N-terminus, a proline-rich region around the first cysteine immediately followed by a strictly conserved glycine residue, and an arginine/lysine-rich region around the second cysteine (Figure 3). These peptides are clearly paralogs and the consensus sequence of the mature peptides is defined as XIXXPXFXPPTCPXGXXXXXXGXCRXXFX.

Interestingly, the three peptides U<sub>17</sub>-Tb1a, U<sub>17</sub>-Tb1b and U<sub>17</sub>-Tb1c exhibited a primary structure with a missing 203 Da mass unit compared to the experimental measurement (Table



**Figure 5.** Minimum evolution phylogenetic trees of peptides precursors in the venom of *T. bicarinatum*. Alignments of preproregions were generated with the Muscle program in Seaview 1:4.5.4.8-2, then analyzed with MEGA 7.0.26 with minimum evolution statistical test and edited using Seaview 1:4.5.4.8-2. Precursors in the venom of *T. bicarinatum* are in colored boxes. (A) Superfamily-A, pilosulin, other venom peptides and antimicrobial peptide precursors. (B) Superfamily-B and antimicrobial peptide precursors. (C) Superfamily-C and secapin precursors.

2). This mass default was consistent with a O-glycosylation by an N-acetyl-hexosamine (HexNac).<sup>24–26</sup> The LC–MS analysis of the three peptides revealed that the difference in mass arose from the N-terminal and more precisely from the common glycosylation site constituted by the N-terminal threonine. The U<sub>17</sub>-Tb1b peptide also has a serine residue at position 11, another possible site of O-glycosylation, though no evidence of such PTM was found on this residue. The nonglycosylated

counterparts of these three peptides were also present in the venom at a slightly lower proportion as Tb-2467, Tb-2570 and Tb-2650 peptides, respectively.

Since the O-glycosylation of peptide toxins has seldom been found in animal venoms,<sup>27</sup> we decided to further investigate the type of O-linked HexNAc in U<sub>17</sub>-Tb1a, U<sub>17</sub>-Tb1b and U<sub>17</sub>-Tb1c by assessing the MS/MS fragmentation of their oxonium ions. Indeed, the MS/MS spectrum yields specific fragments

which enables the identification of the isomeric forms of sugars.<sup>28</sup> In particular, the ions at 138 and 144  $m/z$  are relevant toward discriminating an N-acetylglucosamine (GlcNAc) from an N-acetylgalactosamine (GalNAc) since the intensity ratio of these ions significantly differs for each sugar<sup>29</sup> (Figure 4a). After the experimental validation of this approach using models of N-protected GalNAc- and GlcNAc-threonine, its application to the three glycosylated peptide unambiguously shows a characteristic ratio in favor of glycosylations by a GalNAc (Figure 4b).

### 3.3. Phylogenetic Relationships of Venom Peptide Precursors

The precursors belonging to the superfamily-A exhibited different lengths due to the number of AD or AE motif occurrences in their pro-region. The repetition of these motifs is also observed in the pilosulin-related precursors, as well as in the precursors of melittin and  $\alpha$ -pompilidotoxin.<sup>30,31</sup> The melittin precursor shares 35% identity with the U<sub>3</sub>-MYRTX precursors (prepro-U<sub>3</sub>-Tb1a, -b and -c) while the precursor of  $\alpha$ -pompilidotoxin has 30% sequence identity with the U<sub>4</sub>-Tb1a precursor. It is worth noting that the prepro-U<sub>4</sub>-Tb1a was also related to the drosomycin precursor (46% sequence identity), an antimicrobial peptide found in the hemolymph of *Drosophila melanogaster*.<sup>32</sup> A phylogenetic tree reconstruction from prepro-sequence alignment with minimum evolution (ME) statistical method placed prepro-U<sub>3</sub> sequences in the same cluster as drosomycin and  $\alpha$ -pompilidotoxin (Figure 5A). As expected, other prepro-sequences from superfamily-A precursors fell into the same cluster as pilosulin.

Although prepro-peptides of the superfamily-B had no obvious similarity with other known venom toxin precursors, the consensus prepro-sequence exhibited about 20% identity with prepro-sequences of the antimicrobial *Drosophila* peptides cecropin, adropin, drosomycin and metchnikowin, the maximum shared identity being observed between cecropin and both U<sub>16</sub>-Tb1a and U<sub>13</sub>-Tb1a precursors (40% sequence identity) (Figure 5B). On the ME phylogenetic tree, prepro-U<sub>13</sub>-Tb1a and prepro-U<sub>16</sub>-Tb1a (precursor family B2) fell into the same cluster as cecropin and precursors from the B1 family cluster with metchnikowin (Figure 5B).

The maturation of superfamily-C precursors generated native peptides related to secapins, a group of multifunctional peptides found in the venom of several hymenoptera and even in the hemolymph of the honey-bee as a part of its innate immunity.<sup>33</sup> ADNC encoding secapins have also been characterized in nonvenomous insects such as drosophila (dromepin-3, cDNA accession number JX977165) and mosquitoes (secapin-3, accession number HQ170505, secapin-2a -2b, -2c accession number HQ170502, HQ170503 and HQ170504, respectively). The homologies of *T. bicarinatum* venom peptides with other hymenoptera toxins and, more importantly, with antimicrobial peptides from the hemolymph of insects, support the hypothesis that hymenopteran venom peptides may be related to their innate immune system.

## 4. DISCUSSION

### 4.1. Strategy for Identifying Peptides in the Venom of *T. bicarinatum*

Pioneering work by Bouzid et al.<sup>11</sup> resulted in the first transcriptomic investigation of the venom glands of *Tetramorium bicarinatum*. This work has generated a total of 35 745 putative transcripts which were deposited in the GenBank TSA

database. Although that study permitted the identification of several putative proteins, short gene sequences such as those encoding peptide toxins were largely ignored. It is also worth noting that this transcriptomic analysis only describes hypothetical toxins, since the putative peptides and proteins that were inferred from that study have not been paired with any corresponding masses detected through the mass spectrometry of the venom.

Herein, we describe the large-scale identification of the peptide toxins in the venom of *T. bicarinatum* as well as their expression in the venom glands (Table S1). As mentioned above, our integrative strategy includes three complementary approaches: a proteomic analysis (Edman degradation and/or mass spectrometry-based sequencing), a direct transcriptomic analysis of the venom glands (RT-PCR and DNA sequencing) and the data mining of the available RNA-seq transcriptome. This venomics approach led to the identification of 37 venom peptide precursors which were classified into three superfamilies (-A, -B and -C). A total of 29 mature peptides were confirmed in the venom of *T. bicarinatum* through mass spectrometry.

### 4.2. Evolution of Superfamilies of *T. bicarinatum* Venom Precursors

The conserved prepro-regions and the variability of mature sequences observed in the superfamilies of *T. bicarinatum* venom precursors suggest that venom peptides have evolved from a relatively small set of ancestral genes presumably through duplication followed by the divergence of copies in the regions encoding mature peptides. For instance, the comparison of nucleotide sequences encoding for M, U<sub>1</sub>, U<sub>9</sub>, U<sub>8</sub>-MYRTX-Tb1a precursors showed a weak number of nonsynonymous substitutions (3–10) and codon deletion (5–6) in the prepro-regions while the sequences of mature regions displayed a high rate of mutation (Figure S1). The comparison of encoding sequences for superfamily-B precursors led to the same observation with 1 to 11 nonsynonymous substitutions on 85 nucleotides in the prepro-region and an excess of nonsynonymous substitutions in their mature sequences (Figure S1).

This conservation of prepro-regions and divergence of mature regions has been observed for the antimicrobial peptides<sup>34</sup> found in frog venoms as well as for the venom peptides of spiders and cone snails.<sup>35,36</sup> A high rate of polymorphisms (intra- and interspecific) has also been observed in the antimicrobial peptide (AMP) genes of the *Drosophila* genus.<sup>37,38</sup> Some authors hypothesized that this polymorphism within species is linked to selective forces such as infection pathogen pressure. Ants use venom to capture a broad array of arthropods and also to protect themselves against microbial infection. The divergence in the primary sequence of mature peptides may indicate a range of molecular targets and/or mode of action. Thus, we hypothesized that differential exposure to prey and microbial pathogens has imposed selective pressures toward toxin evolution leading to a high rate of mutations in the mature sequences.

The superfamily-C precursors were related to secapins. Since the description of the first secapin in honey-bee venom,<sup>39</sup> several other secapin peptides have been reported in bees, ants, wasps, drosophila, mosquitoes and spiders. Surprisingly, the prepro-sequence of *T. bicarinatum* venom secapins shares little identity with other secapin precursors including secapins described from other ant species. On the ME phylogenetic tree

generated from the prepro-sequences of all characterized arthropod secapins, those of *T. bicarinatum* venom cluster separately (Figure 5C). However, the BLAST of mature secapin sequences on the transcriptome of *T. bicarinatum* allowed us to predict two other secapin precursors known as prepro-Tb3587 and prepro-Tb2681 (contigs GASM01032508.1 and GASM01033385.1). Their prepro-sequences cluster on the ME phylogenetic tree with wasp navisecapin (ADT89766.1, *Nasonia vitripennis*) and ant soinpin-3 (AGCS0820.1, *Solenopsis invicta*). Moreover, the BLAST of the prepro-sequences of the *T. bicarinatum* superfamily-C on the transcriptome assemblies available in the NCBI database has led to the identification of multiple novel secapin precursors in the Formicidae. Thus, 16 novel secapin precursors were found in both the Myrmicinae and the Formicinae subfamilies and share about 53% identity with the secapin consensus prepro-sequence of *T. bicarinatum* (Figure S2).

In contrast, the mature region of secapin precursors are well conserved in insects and they share about 34% sequence identity with the consensus sequence of mature *T. bicarinatum* secapins. This evolutive pressure on the mature sequence but not on the signal and pro-regions is often observed. For instance, this has been reported in vertebrates for the RFamide 26RFa/QRFP neuropeptide whose the C-terminal extremity which ensures certain biological activities is particularly well conserved.<sup>40,41</sup> The conservation of the mature primary sequences PPTCPXG and GXCR may highlight their functional importance in secapin folding and/or target recognition. The diversity of prepro-regions in addition to the conservation of mature sequences of secapins lead to hypothesize that convergence mechanisms drive the evolution of the genes encoding the different clusters of secapin precursors.

**4.2.1. Presumed Biological Functions of Mature Peptides.** The superfamilies -A and -B of precursors encoded for a set of peptides that displayed a remarkably low sequence identity. Despite this variety, several peptides shared physicochemical properties characteristic of linear AMPs such as amphipathy, polycationicity and  $\alpha$ -helix propensity and most are likely membrane-active peptides. The physicochemical properties of AMPs, which do not require a strictly conserved pattern of amino acids to be functional, may explain the high diversity of toxin sequences observed in the venom peptidome of *T. bicarinatum*. Hymenopteran venoms are known to be rich in linear AMPs. For instance, M-poneritoxins-Ng (ponericins),<sup>42</sup> M-poneritoxins-Dq (dinoponeratoxins),<sup>17</sup> M-myrmecitoxin-Mp1a (pilosulin 1),<sup>43</sup> M-ectatotoxin-Eb2a-c (ponericin-Q42, -Q49 and -Q50)<sup>44</sup> and M-MYRTX-Tb1a (bicarinalin), which is the most abundant peptide in the *T. bicarinatum* venom,<sup>14</sup> have been characterized as ant AMPs. Furthermore, some peptides described in the present manuscript (i.e., U<sub>10</sub>-Tb1a and U<sub>12</sub>-Tb1a) exhibited homologies with antimicrobial and insecticidal poneritoxins-Ng (ponericins) and antimicrobial solitary wasp peptides (VP4a, VP4b).<sup>45</sup> Altogether, our data suggest that most of the peptides yielded by the superfamilies -A and -B were polycationic and amphiphilic peptides with likely antimicrobial, cytotoxic and insecticidal properties.

The mature peptides belonging to the superfamily-A of precursors predominated in the *T. bicarinatum* venom at more than 74% of the peptidome content and, thus, highlighted the important functional roles of these myrmecitoxins. In the

superfamily-A, the peptides from the three U<sub>3</sub>-myrmecitoxins plus their cleaved version represented around 25% of the whole peptidome. The three full-length peptides U<sub>3</sub>-MYRTX-Tb1a, -b and -c were markedly different from other peptides of the peptidome in being globally hydrophobic with a positively charged C-terminal extremity (Table 2 and Table S2). They displayed sequence homologies with the peptides  $\delta$ -PPONTX-Pc1a (poneratoxin) in the venom of the ant *Paraponera clavata* (Figure 3).  $\delta$ -PPONTX-Pc1a is a neurotoxic peptide affecting the sodium channels in both vertebrates and invertebrates.<sup>46</sup> These data suggest that the U<sub>3</sub>-MYRTX group of peptides may play an important role in the paralysis of prey probably by affecting sodium channels.

The secapin peptides in the venom of hymenoptera are known to display several activities including neurotoxic, hyperalgesic, edematogenic, antifibrinolytic, antielastolytic and antibacterial properties.<sup>33,47–49</sup> Secapin-like peptides from the superfamily-C of precursors might have similar functions. Interestingly, some secapin-like peptides in the venom of *T. bicarinatum* exhibited the O-glycosylation of the N-terminal threonine residue. To the best of our knowledge, such PTM has not been previously reported for ant venom peptides nor for other secapin peptides. Nevertheless, O-glycosylated AMPs, named formaecins, have been isolated in the hemolymph of the ant *Myrmecia gulosa* and were significantly more active than their nonglycosylated counterparts.<sup>50</sup> Interestingly, several O-glycosylated peptides have been identified in the venom of cone snails,<sup>51</sup> scorpions,<sup>52</sup> snakes,<sup>27</sup> and wasps.<sup>53</sup> None of these other glycosylated peptides exhibit any significant sequence homology with the glycosylated myrmecitoxins described in the present study. The O-glycosylation is generally important for biological activity and may enhance the potency and the stability of venom peptides. Nevertheless, further studies on these three glycosylated myrmecitoxins (U<sub>17</sub>-Tb1a, U<sub>17</sub>-Tb1b and U<sub>17</sub>-Tb1c) are needed to determine the biological role of this glycan.

#### 4.3. Peptide Fragmentation and Degradation in Ant Venom

Recent studies using highly sensitive mass spectrometry reveal that ant venoms can contain several thousands of peptides. In this study, we have detected more than 2800 peptide masses in the venom of *T. bicarinatum*. However, each signal does not reflect the presence of an individual peptide in the venom. Indeed, several signals can be assigned to the same peptide like sodium and potassium adducts or correspond to artifactual species such as intermediate fragments or peptides without certain post translational modifications lost during the experimental processes. Nevertheless, the discrepancy between the masses detected by Orbitrap MS and those detected by the LCQ Advantage cannot be solely explained by adducts and artifacts and suggests that hundreds of peptides are present in the venom. Dutertre et al. (2013) showed that thousands of peptides arise from a limited number of gene transcripts (~100) in the venom of the cone snail *Conus marmoreus* due to variable N- and C-terminal truncations and alternative cleavage sites.<sup>22</sup> Additionally, Kazuma et al. (2017) and Cologna et al. (2018) found in the venoms of both ants *Odontomachus monticola* and *Neoponera villosa* intact pilosulin-like peptides and ponericin-like peptides plus several truncated forms reinforcing the idea that mature peptides can be cleaved in the venom.<sup>9,54</sup> The *de novo* mass spectrometry of the venom

of *T. bicarinatum* indicates that similar peptide processing occurred. This is exemplified by the precursor prepro-U<sub>1</sub>-MYRTX-Tb1a which generates at least 32 peptides in the venom including part of the pro-peptide (Table S4). The pro-peptides found in this venom are likely to be artifactual to the maturation process or come from the content of venom gland cells. Other examples from our study are the peptide fragments found in the U<sub>3</sub>-MYRTX group, the U<sub>10</sub>-MYRTX group and the U<sub>13</sub>-MYRTX group which seem to be preferentially cleaved around a methionine residue. However, it remains to be known if these fragmented peptides, which are likely related to peptide degradation, are functionally active.

## 5. CONCLUDING REMARKS

In summary, we provide the first insights into the entire peptide arsenal contained in an ant venom. The integrated methodology employed in this study allowed the identification of 37 venom peptide precursors, which are processed into several thousand peptides in the venom. Indeed, our mass spectrometry analysis revealed that peptide fragmentation and variable N- and C-terminal truncations generated an enormous number of myrmicitoxins from a limited set of precursors. We also demonstrated that three myrmicitoxins were O-glycosylated on their N-terminal threonine residue. Most of the myrmicitoxins found in *T. bicarinatum* venom are presumably membrane-active peptides and may possess antimicrobial, cytolytic and insecticidal properties. It is also worth noting that *T. bicarinatum* venom glands expressed homologous toxins of the neurotoxic peptide δ-PPONTX-Pc1a (poneratoxin) only known in the venom of the ant *Paraponera clavata*. Nevertheless, future functional studies are required to confirm the biological activities of these myrmicitoxins. From an evolutionary perspective, our data suggest that several of the myrmicitoxins described here may have evolved from the innate immunity system of ants.

## ■ ASSOCIATED CONTENT

### S Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jproteome.8b00452](https://doi.org/10.1021/acs.jproteome.8b00452).

Figures S1–S2, Tables S1–S4 ([PDF](#))

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Phone: +(33) 5 63 48 64 32. E-mail: [axel.touchard2@gmail.com](mailto:axel.touchard2@gmail.com).

### ORCID

Axel Touchard: [0000-0002-7766-0088](https://orcid.org/0000-0002-7766-0088)

Jérôme Leprince: [0000-0002-7814-9927](https://orcid.org/0000-0002-7814-9927)

### Author Contributions

#Michel Treilhou and Elsa Bonnafé contributed equally.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

PISSARO Platform is cofunded by The European Union and Région Normandie through the European Regional Development Fund (ERDF). We are grateful to Dr. Jérôme Orivel for providing the ant colonies and Andrea Yockey-Dejean for

proofreading the manuscript. We would also like to thank Guilhem Vazzoler for his assistance in the laboratory during his internship.

## ■ REFERENCES

- (1) Casewell, N. R.; Wüster, W.; Vonk, F. J.; Harrison, R. A.; Fry, B. G. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol Evol* **2013**, *28* (4), 219–229.
- (2) King, G. F. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opin. Biol. Ther.* **2011**, *11*, 1469–1484.
- (3) Smith, J.; Herzig, V.; King, G.; Alewood, P. The insecticidal potential of venom peptides. *Cell. Mol. Life Sci.* **2013**, *70* (19), 3665–3693.
- (4) von Reumont, B.; Campbell, L.; Jenner, R. Quo Vadis Venomics? A roadmap to neglected venomous invertebrates. *Toxins* **2014**, *6* (12), 3488–3551.
- (5) Touchard, A.; Aili, S. R.; Fox, E. G. P.; Escoubas, P.; Orivel, J.; Nicholson, G. M.; Dejean, A. The biochemical toxin arsenal from ant venoms. *Toxins* **2016**, *8* (1), 30.
- (6) Touchard, A.; Koh, J. M. S.; Aili, S. R.; Dejean, A.; Nicholson, G. M.; Orivel, J.; Escoubas, P. The complexity and structural diversity of ant venom peptidomes is revealed by mass spectrometry profiling. *Rapid Commun. Mass Spectrom.* **2015**, *29*, 385–396.
- (7) Aili, S. R.; Touchard, A.; Koh, J. M. S.; Dejean, A.; Orivel, J.; Padula, M. P.; Escoubas, P.; Nicholson, G. M. Comparisons of protein and peptide complexity in poneroid and formicoid ant venoms. *J. Proteome Res.* **2016**, *15*, 3039.
- (8) Aili, S. R.; Touchard, A.; Escoubas, P.; Padula, M. P.; Orivel, J.; Dejean, A.; Nicholson, G. M. Diversity of peptide toxins from stinging ant venoms. *Toxicon* **2014**, *92*, 166–178.
- (9) Kazuma, K.; Masuko, K.; Konno, K.; Inagaki, H. Combined venom gland transcriptomic and venom peptidomic analysis of the predatory ant *Odontomachus monticola*. *Toxins* **2017**, *9* (10), 323.
- (10) Torres, A. F. C.; Huang, C.; Chong, C. M.; Leung, S. W.; Prieto-da-Silva, Á. R. B.; Havit, A.; Quinet, Y. P.; Martins, A. M. C.; Lee, S. M. Y.; Rádis-Baptista, G. Transcriptome analysis in venom gland of the predatory giant ant *Dinoponera quadriceps*: Insights into the polypeptide toxin arsenal of hymenopterans. *PLoS One* **2014**, *9* (1), e87556.
- (11) Bouzid, W.; Verdenaud, M.; Klopp, C.; Ducancel, F.; Noirot, C.; Vétillard, A. De Novo sequencing and transcriptome analysis for *Tetramorium bicarinatum*: a comprehensive venom gland transcriptome analysis from an ant species. *BMC Genomics* **2014**, *15* (1), 987.
- (12) Escoubas, P.; King, G. F. Venomics as a drug discovery platform. *Expert Rev. Proteomics* **2009**, *6*, 221.
- (13) Escoubas, P.; Quinton, L.; Nicholson, G. M. Venomics: unravelling the complexity of animal venoms with mass spectrometry. *J. Mass Spectrom.* **2008**, *43* (3), 279–295.
- (14) Rifflet, A.; Gavalda, S.; Téné, N.; Orivel, J.; Leprince, J.; Guilhaudis, L.; Génin, E.; Vétillard, A.; Treilhou, M. Identification and characterization of a novel antimicrobial peptide from the venom of the ant *Tetramorium bicarinatum*. *Peptides* **2012**, *38* (2), 363–370.
- (15) Bouzid, W.; Klopp, C.; Verdenaud, M.; Ducancel, F.; Vétillard, A. Profiling the venom gland transcriptome of *Tetramorium bicarinatum* (Hymenoptera: Formicidae): The first transcriptome analysis of an ant species. *Toxicon* **2013**, *70* (0), 70–81.
- (16) Téné, N.; Bonnafé, E.; Berger, F.; Rifflet, A.; Guilhaudis, L.; Ségalas-Milazzo, I.; Pipy, B.; Coste, A.; Leprince, J.; Treilhou, M. Biochemical and biophysical combined study of bicarinatin, an ant venom antimicrobial peptide. *Peptides* **2016**, *79*, 103–113.
- (17) Johnson, S. R.; Copello, J. A.; Evans, M. S.; Suarez, A. V. A biochemical characterization of the major peptides from the venom of the giant Neotropical hunting ant *Dinoponera australis*. *Toxicon* **2010**, *55* (4), 702–710.

- (18) King, G. F.; Gentz, M. C.; Escoubas, P.; Nicholson, G. M. A rational nomenclature for naming peptide toxins from spiders and other venomous animals. *Toxicon* **2008**, *52* (2), 264–276.
- (19) Touchard, A.; Dauvois, M.; Arguel, M.-J.; Petitclerc, F.; Leblanc, M.; Dejean, A.; Orivel, J.; Nicholson, G. M.; Escoubas, P. Elucidation of the unexplored biodiversity of ant venom peptidomes via MALDI–TOF mass spectrometry and its application for chemotaxonomy. *J. Proteomics* **2014**, *105*, 217–231.
- (20) Cologna, C. T.; Cardoso, J. d. S.; Jourdan, E.; Degueldre, M.; Upert, G.; Gilles, N.; Uetanabaro, A. P. T.; Costa Neto, E. M.; Thonart, P.; de Pauw, E. Peptidomic comparison and characterization of the major components of the venom of the giant ant *Dinoponera quadriceps* collected in four different areas of Brazil. *J. Proteomics* **2013**, *94*, 413–422.
- (21) Escoubas, P.; Sollod, B.; King, G. F. Venom landscapes: mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. *Toxicon* **2006**, *47* (6), 650–63.
- (22) Dutertre, S.; Jin, A. H.; Kaas, Q.; Jones, A.; Alewood, P. F.; Lewis, R. J. Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol. Cell. Proteomics* **2013**, *12* (2), 312–29.
- (23) Inagaki, H.; Akagi, M.; Imai, H. T.; Taylor, R. W.; Kubo, T. Molecular cloning and biological characterization of novel antimicrobial peptides, pilosulin 3 and pilosulin 4, from a species of the Australian ant genus *Myrmecia*. *Arch. Biochem. Biophys.* **2004**, *428* (2), 170–178.
- (24) Wada, Y.; Tajiri, M.; Yoshida, S. Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. *Anal. Chem.* **2004**, *76* (22), 6560–6565.
- (25) Nwosu, C. C.; Seipert, R. R.; Strum, J. S.; Hua, S. S.; An, H. J.; Zivkovic, A. M.; German, B. J.; Lebrilla, C. B. Simultaneous and extensive site-specific N-and O-glycosylation analysis in protein mixtures. *J. Proteome Res.* **2011**, *10* (5), 2612–2624.
- (26) Halim, A.; Brinkmalm, G.; Rüetschi, U.; Westman-Brinkmalm, A.; Portelius, E.; Zetterberg, H.; Blennow, K.; Larson, G.; Nilsson, J. Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein/amyloid  $\beta$ -peptides in human cerebrospinal fluid. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (29), 11848–11853.
- (27) Degueldre, M.; Echterbille, J.; Smargiasso, N.; Damblon, C.; Gouin, C.; Mourier, G.; Gilles, N.; De Pauw, E.; Quinton, L. In-depth glyco-peptidomics approach reveals unexpected diversity of glycosylated peptides and atypical post-translational modifications in *Dendroaspis angusticeps* snake venom. *Int. J. Mol. Sci.* **2017**, *18* (11), 2453.
- (28) Treilhou, M.; Ferro, M.; Monteiro, C.; Poinsot, V.; Jabbouri, S.; Kanony, C.; Promé, D.; Promé, J.-C. Differentiation of O-acetyl and O-carbamoyl esters of N-acetyl-glucosamine by decomposition of their oxonium ions. Application to the structure of the nonreducing terminal residue of Nod factors. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (4), 301–311.
- (29) Halim, A.; Westerlind, U.; Pett, C.; Schorlemer, M.; Rüetschi, U.; Brinkmalm, G.; Sihlbom, C.; Lengqvist, J.; Larson, G. r.; Nilsson, J. Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC–MS/MS of glycopeptides. *J. Proteome Res.* **2014**, *13* (12), 6024–6032.
- (30) Vlasak, R.; Unger-ullmann, C.; Kreil, G.; Frischaufl, A. M. Nucleotide sequence of cloned cDNA coding for honeybee prepromelittin. *Eur. J. Biochem.* **1983**, *135* (1), 123–126.
- (31) Konno, K.; Hisada, M.; Itagaki, Y.; Naoki, H.; Kawai, N.; Miwa, A.; Yasuhara, T.; Takayama, H. Isolation and Structure of Pompilidotoxins, Novel Peptide Neurotoxins in Solitary Wasp Venoms. *Biochem. Biophys. Res. Commun.* **1998**, *250* (3), 612–616.
- (32) Cohen, L.; Moran, Y.; Sharon, A.; Segal, D.; Gordon, D.; Gurevitz, M. Drosomycin, an innate immunity peptide of *Drosophila melanogaster*, interacts with the fly voltage-gated sodium channel. *J. Biol. Chem.* **2009**, *284* (35), 23558–23563.
- (33) Lee, K. S.; Kim, B. Y.; Yoon, H. J.; Choi, Y. S.; Jin, B. R. Secapin, a bee venom peptide, exhibits anti-fibrinolytic, anti-elastolytic, and anti-microbial activities. *Dev. Comp. Immunol.* **2016**, *63*, 27–35.
- (34) Vanhoye, D.; Bruston, F.; Nicolas, P.; Amiche, M. Antimicrobial peptides from hylid and ranin frogs originated from a 150-million-year-old ancestral precursor with a conserved signal peptide but a hypermutable antimicrobial domain. *Eur. J. Biochem.* **2003**, *270* (9), 2068–2081.
- (35) Olivera, B. M.; Walker, C.; Cartier, G. E.; Hooper, D.; Santos, A. D.; Schoenfeld, R.; Shetty, R.; Watkins, M.; Bandyopadhyay, P.; Hillyard, D. R. Speciation of cone snails and interspecific hyperdivergence of their venom peptides: potential evolutionary significance of introns. *Ann. N. Y. Acad. Sci.* **1999**, *870* (1), 223–237.
- (36) Tang, X.; Zhang, Y.; Hu, W.; Xu, D.; Tao, H.; Yang, X.; Li, Y.; Jiang, L.; Liang, S. Molecular diversification of peptide toxins from the tarantula *Haplopelma hainanum* (*Ornithoctonus hainana*) venom based on transcriptomic, peptidomic, and genomic analyses. *J. Proteome Res.* **2010**, *9* (5), 2550–2564.
- (37) Hanson, M. A.; Hamilton, P. T.; Perlman, S. J. Immune genes and divergent antimicrobial peptides in flies of the subgenus *Drosophila*. *BMC Evol. Biol.* **2016**, *16* (1), 228.
- (38) Unckless, R. L.; Lazzaro, B. P. The potential for adaptive maintenance of diversity in insect antimicrobial peptides. *Philos. Trans. R. Soc., B* **2016**, *371* (1695), 20150291.
- (39) Gauldie, J.; Hanson, J. M.; Rumjanek, F. D.; Shipolini, R. A.; Vernon, C. A. The peptide components of bee venom. *Eur. J. Biochem.* **1976**, *61* (2), 369–376.
- (40) Liu, Y.; Zhang, Y.; Li, S.; Huang, W.; Liu, X.; Lu, D.; Meng, Z.; Lin, H. Molecular cloning and functional characterization of the first non-mammalian 26RFa/QRFP orthologue in goldfish, *Carassius auratus*. *Mol. Cell. Endocrinol.* **2009**, *303* (1), 82–90.
- (41) Leprince, J.; Bagnol, D.; Bureau, R.; Fuksumi, S.; Granata, R.; Hinuma, S.; Larhammar, D.; Primeaux, S.; Sopkova-de Oliveira Santos, J.; Tsutsui, K.; Ukena, K.; Vaudry, H. The Arg–Phe-amide peptide 26RFa/glutamine RF-amide peptide and its receptor: IUPHAR Review 24. *Br. J. Pharmacol.* **2017**, *174* (20), 3573–3607.
- (42) Orivel, J.; Redeker, V.; Le Caer, J. P.; Krier, F.; Revol-Junelles, A. M.; Longeon, A.; Chaffotte, A.; Dejean, A.; Rossier, J. Ponericins, new antibacterial and insecticidal peptides from the venom of the ant *Pachycondyla goeldii*. *J. Biol. Chem.* **2001**, *276* (21), 17823–9.
- (43) Donovan, G. R.; Baldo, B. A.; Sutherland, S. Molecular cloning and characterization of a major allergen (Myr p I) from the venom of the Australian jumper ant, *Myrmecia pilosula*. *Biochim. Biophys. Acta, Gene Struct. Expression* **1993**, *1171* (3), 272–80.
- (44) Pluzhnikov, K. A.; Kozlov, S. A.; Vassilevski, A. A.; Vorontsova, O. V.; Feofanov, A. V.; Grishin, E. V. Linear antimicrobial peptides from *Ectatomma quadridens* ant venom. *Biochimie* **2014**, *107*, 211–215.
- (45) Baek, J. H.; Lee, S. H. Differential gene expression profiles in the venom gland/sac of *Eumenes pomiformis* (Hymenoptera: Eumenidae). *Toxicon* **2010**, *55* (6), 1147–1156.
- (46) Piek, T.; Duval, A.; Hue, B.; Karst, H.; Lapiède, B.; Mantel, P.; Nakajima, T.; Pelhate, M.; Schmidt, J. O. Poneratoxin, a novel peptide neurotoxin from the venom of the ant, *Paraponera clavata*. *Comp. Biochem. Physiol., C: Comp. Pharmacol.* **1991**, *99* (3), 487–495.
- (47) Mourelle, D.; Brigatte, P.; Bringanti, L.; De Souza, B.; Arcuri, H.; Gomes, P.; Baptista-Saidemberg, N.; Neto, J. R.; Palma, M. Hyperalgesic and edematogenic effects of Secapin-2, a peptide isolated from Africanized honeybee (*Apis mellifera*) venom. *Peptides* **2014**, *59*, 42–52.
- (48) Vlasak, R.; Kreil, G. Nucleotide sequence of cloned cDNAs coding for preprosecapin, a major product of queen-bee venom glands. *Eur. J. Biochem.* **1984**, *145* (2), 279–282.
- (49) Hou, C.; Guo, L.; Lin, J.; You, L.; Wu, W. Production of antibacterial peptide from bee venom via a new strategy for heterologous expression. *Mol. Biol. Rep.* **2014**, *41* (12), 8081–8091.
- (50) Mackintosh, J. A.; Veal, D. A.; Beattie, A. J.; Gooley, A. A. Isolation from an ant *Myrmecia gulosa* of two inducible O-glycosylated proline-rich antibacterial peptides. *J. Biol. Chem.* **1998**, *273* (11), 6139–43.

- (51) Craig, A. G.; Norberg, T.; Griffin, D.; Hoeger, C.; Akhtar, M.; Schmidt, K.; Low, W.; Dykert, J.; Richelson, E.; Navarro, V.; Mazella, J.; Watkins, M.; Hillyard, D.; Imperial, J.; Cruz, L. J.; Olivera, B. M. Contulakin-G, an O-Glycosylated Invertebrate Neurotensin. *J. Biol. Chem.* **1999**, *274* (20), 13752–13759.
- (52) Hassani, O.; Loew, D.; Van Dorsselaer, A.; Papandréou, M. J.; Sorokine, O.; Rochat, H.; Sampieri, F.; Mansuelle, P. Aah VI, a novel, N-glycosylated anti-insect toxin from *Androctonus australis hector* scorpion venom: isolation, characterisation, and glycan structure determination. *FEBS Lett.* **1999**, *443* (2), 175–180.
- (53) Piek, T.; Hue, B.; Le Corronc, H.; Mantel, P.; Gobbo, M.; Rocchi, R. Presynaptic block of transmission in the insect CNS by mono-and di-galactosyl analogues of vespulakinin 1, a wasp (*Paravespula maculifrons*) venom neurotoxin. *Comp. Biochem. Physiol., C: Comp. Pharmacol.* **1993**, *105* (2), 189–196.
- (54) Cologna, C. T.; Rodrigues, R. S.; Santos, J.; de Pauw, E.; Arantes, E. C. Quinton, L. Peptidomic investigation of *Neoponera villosa* venom by high-resolution mass spectrometry: seasonal and nesting habitat variations. *J. Venomous Anim. Toxins Incl. Trop. Dis.* **2018**, *24* (1), 6.